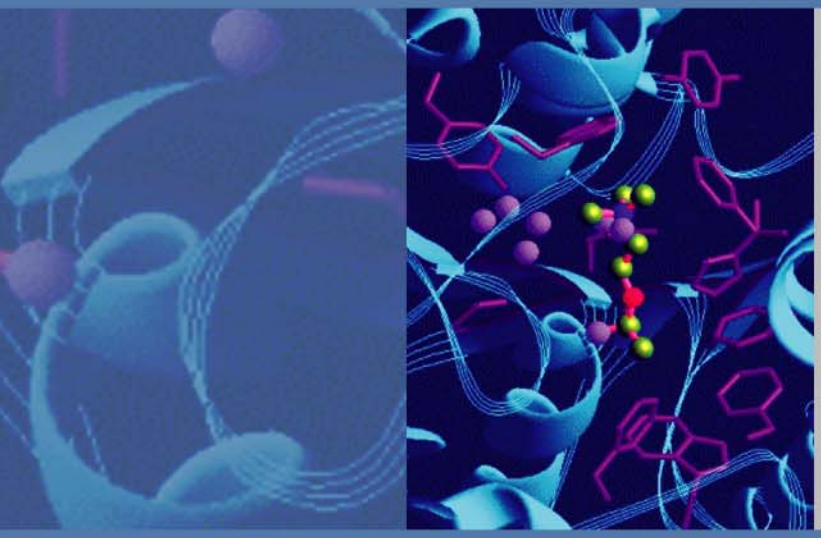


Xcalibur

Proteome Discoverer

Version 1.3

User Guide

XCALI-97358 Revision A May 2011

DOCUMENTATION SURVEY

Part of Thermo Fisher Scientific

Thermo SCIENTIFIC

© 2011 Thermo Fisher Scientific Inc. All rights reserved.

Xcalibur is a registered trademark of Thermo Fisher Scientific Inc. in the United States. ZCore and Proteome Discoverer are trademarks of Thermo Fisher Scientific Inc. in the United States.

SEQUEST is a registered trademark of the University of Washington in the United States.

iTRAQ is a registered trademark of Applera Corporation in the United States and possibly other countries.

TMT is a registered trademark of Proteome Sciences plc in the United Kingdom.

Mascot is a registered service mark of Matrix Science Ltd. in the United States.

Excel, Microsoft, and Windows are registered trademarks of Microsoft Corporation in the United States and other countries.

All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.

Thermo Fisher Scientific Inc. provides this document to its customers with a product purchase to use in the product operation. This document is copyright protected and any reproduction of the whole or any part of this document is strictly prohibited, except with the written authorization of Thermo Fisher Scientific Inc.

The contents of this document are subject to change without notice. All technical information in this document is for reference purposes only. System configurations and specifications in this document supersede all previous information received by the purchaser.

Thermo Fisher Scientific Inc. makes no representations that this document is complete, accurate or error-free and assumes no responsibility and will not be liable for any errors, omissions, damage or loss that might result from any use of this document, even if the information in the document is followed properly.

This document is not part of any sales contract between Thermo Fisher Scientific Inc. and a purchaser. This document shall in no way govern or modify any Terms and Conditions of Sale, which Terms and Conditions of Sale shall govern all conflicting information between the two documents.

Release history: Revision A, April, 2011

Software version: Thermo Proteome Discoverer version 1.3

For Research Use Only. Not for use in diagnostic procedures.

Contents

Preface . . . . . **xv**
Related Documentation xv
System Requirements xvi
Special Notices xvii
Contacting Us xvii

Chapter 1 Introduction **1**
Features 1
Search Engines 3
Wizards and Workflow Editor 5
Quantification 6
The Qual Browser Application 7
Peptides and Fragment Ions 7
Fragmentation Methods 8
MudPIT Experiments 9
Workflow 9
Inputs and Outputs 12
FASTA Databases 12
Inputs 12
Outputs 13
Limitations 14
New Features in This Release 14
Protein Annotation 14
Data Display 14
PTM Analysis 15
Data Exportation 15
Protein Grouping 15
Peptide Grouping 16
Peptide Confidence 16
Workflow Editor Nodes 16
Fragment Matching 17
64-Bit Version 17
Manual Creation of Decoy Databases 17

Chapter 2 Using the Proteome Discoverer Interface 19
Proteome Discoverer User Interface. 19
Main Window. 19
Colors 21
Menu Bar 22
Shortcut Menus. 23
Toolbar 24
Status Bar 27
Job Queue 28
Search Results Window Icons 33
Dialog Box Icons. 35
Customizing the User Interface 36
Customizing the Toolbar. 36
Customizing Menus 43
Customizing the Job Queue Toolbar. 51
Setting Global Default Fragment Match Options 53

Chapter 3 Getting Started. 61
Starting the Proteome Discoverer Application. 61
Closing the Proteome Discoverer Application 62
Starting a New Search by Using the Search Wizards 63
Configuring Search Parameters 64
Starting a New Search 71
Search Wizard Parameters 83
Starting a New Search by Using the Workflow Editor. 89
Workflow Editor Parameters 91
Creating a Search Workflow 91
Creating a Search Workflow for Multiple Raw Files from the Same Sample 101
Creating a Quantification Workflow. 103
Creating an Annotation Workflow 103
Creating a PTM Analysis Workflow 103
Creating Parallel Workflows 105
Adding a Non-Fragment Filter Node for High-Resolution Data 106
Opening an Existing Workflow. 108
Deleting an Existing Workflow Template 112
Changing the Name and Description of a Workflow Template. 113
Importing Raw Data Files in Other Formats into a Workflow. 114
Saving a Workflow as an XML Template 114
Exporting Spectra 114
Workflow Nodes. 117

Chapter 4 Using the Proteome Discoverer Daemon Utility 121
Starting the Proteome Discoverer Daemon Application in a Window 121
Selecting the Server 122
Starting a Workflow 123
Creating a Parameter File for Use with the Xcalibur Data System 125
Monitoring Job Execution in the Proteome Discoverer Daemon Application 127
Proteome Discoverer Daemon Application Window Parameters 128
Logging on to a Remote Server 130
Running the Proteome Discoverer Daemon Application from the Xcalibur Data System 132
Before You Start 132
Batch Processing Using a Processing Method 134
Setting Up and Configuring the Mass Spectrometer 136
Setting Up the Experiment Method with the Sequence Editor 136
Selecting a Single Sample. 139
Batch Processing by Using a Post-Acquisition Method 144
Processing MudPIT Samples by Using a Processing Method 149
MudPIT Processing Using the Run Sequence Dialog Box 153
Running the Proteome Discoverer Daemon Application on the Command Line 153
Syntax 153
Examples 156

Chapter 5 Searching for Data 159
Using FASTA Databases 159
Displaying FASTA Files 159
Adding FASTA Files 162
Deleting FASTA Files 162
Compressing a Protein Database 163
Displaying Temporary FASTA Files 164
Adding a Protein Sequence and Reference to a FASTA Database File 164
Finding Protein Sequences and References 165
Compiling a FASTA Database. 170
Excluding Individual Protein References and Sequences from a FASTA Database. 174
FASTA Database Utilities Dialog Box Parameters 175
Enter Filter Criteria for Reference Dialog Box Parameters 179
Managing FASTA Indexes. 179

Chemical Modifications 195
Dynamic Modifications 196
Static Modifications 196
Opening the Chemical Modifications View 196
Adding Chemical Modifications 200
Adding Amino Acids 201
Deleting Chemical Modifications 202
Importing Chemical Modifications 202
Deleting Amino Acids 204
Using the Qual Browser Application 205
Customizing Cleavage Reagents 206
Cleavage Reagents View Parameters 207
Adding a Cleavage Reagent 211
Deleting a Cleavage Reagent 211
Modifying a Cleavage Reagent 211
Filtering Cleavage Reagent Data 212

Chapter 6 Interpreting Search Results 213
Using the Results Report 213
Opening MSF Files 215
Saving MSF Files 221
Closing MSF Files 222
Importing SRF Files 222
Organizing Reports 222
Controlling the Search Report Windows 236
Using the Search Report Menu 237
Using the Proteins Page 237
Displaying the Proteins Page 238
Interpreting Your Results with the Protein Identification Details View 245
Interpreting Your Results with the Report Item Distribution Chart 257
Interpreting Your Results with the Sequence Comparison View 283
Interpreting Your Results with the Protein Group Members View 285
Interpreting Your Results with the Protein References of a Peptide View 287

Using the Peptides Page 288
Displaying the Peptides Page 288
Interpreting Your Results with the Peptide Identification Details View 299
Interpreting Your Results with the Report Item Distribution Chart 315
Interpreting Your Results with the Peptide Consensus View 316
Interpreting Your Results with the Dual-Peptide Consensus View. 324
Displaying Post-Translational Modification Information in the UNIMOD Web View 327
Interpreting Your Results with the Chromatogram View. 330
Interpreting Your Results with the Spectrum View 333
Interpreting Your Results with the Fragment Match Spectrum View. 335
Interpreting Your Results with the Precursor Isotope Pattern View 337
Interpreting Your Results with the Extracted Ion Chromatogram View. . . . 340
Interpreting Your Results with the Peptide Group Members View 343
Calculating the Delta Score 345
Using the Search Input Page 346
Search Input Page Parameters 347
Using the Result Filters Page 349
Using the Peptide Confidence Page 349
Using the Search Summary Page 349
Using the Quantification Summary Page. 351
Using the Quan Spectra Page 351
Creating a Report from Multiple Results 351
Finding Common or Unique Proteins in Multiple Searches 353
Applying Filters Specific to Different Searches in Multiconsensus Reports. 356

Chapter 7 Filtering Data 359
Result Filters Page 359
Displaying the Result Filters Page 360
Result Filters Page Parameters 361
Protein Filters 364
Peptide Filters 365
Filtering Data 379
Filtering Results with the Filters on the Result Filters Page 379
Using Filter Sets 389
Removing and Deactivating Filters 392
Filtering Results with Row Filters 393
Displaying Filtered-Out Rows 400

Grouping Proteins 401
Protein Grouping Algorithm 407
Proteins Containing Peptides with Sequences Not Belonging to Master Protein 411
Protein Groups in the Status Bar 412
Proteins Grouped by the Grouping Algorithm in Previous Releases. 412
Number of Unique Peptides Column on Proteins Page 412
PSMs Identified by Multiple Workflow Nodes 412
Grouping Peptides 413
Calculating False Discovery Rates 415
Target False Discovery Rates (FDRs) 416
Peptide Confidence Indicators 416
Setting Up False Discovery Rates in Search Wizards and the Workflow Editor 417
Viewing the Results on the Peptide Confidence Page 424
Recalculating the False Discovery Rates 426
Changing the Target Rate and Filter Settings 427
Peptide Confidence Page Parameters 429

Chapter 8 Protein Annotation 433
ProteinCenter 433
Gene Ontology (GO) Annotation 434
Pfam Annotation 435
Configuring Protein Annotation 435
Creating a Protein Annotation Workflow 437
Displaying the Annotated Protein Results 438
Reannotating MSF Files 443
Uploading Results to ProteinCenter 446
Accessing ProteinCards 449
ProteinCard Parameters 450
General Page 450
Keys Page 452
Features Page 453
Molecular Functions Page 454
Cellular Components Page 455
Biological Processes Page 456
Diseases Page 458
External Links Page 458
GO Slim Categories 459
GO Slim Categories for Molecular Functions 459
GO Slim Categories for Cellular Components 460
GO Slim Categories for Biological Processes 463

Chapter 9 Quantification 467
Activating the Quantification Menu 468
Proteins Included in the Quantification 468
Performing Precursor Ion Quantification 469
SILAC 2plex Methods 469
SILAC 3plex Methods 472
Dimethylation 3plex Method 472
18O Labeling Method 472
Creating a Workflow for Precursor Ion Quantification 472
Performing Reporter Ion Quantification 476
TMT Quantification 476
iTRAQ Quantification 477
Creating a Workflow for Reporter Ion Quantification 477
Performing TMT Quantification on HCD and CID Scans 481
Demonstrating How to Create a Workflow for Reporter Ion Quantification 482
Performing Peak Area Calculation Quantification 483
Setting Up the Quantification Method 485
Specifying the Quantification Channels 488
Setting Up Quantification Channels for Ratio Reporting 495
Setting Up the Ratio Calculation 497
Setting Peptide Parameters Used to Calculate Protein Ratios 500
Correcting Experimental Bias 501
Checking the Quantification Method 502
Restoring Quantification Method Template Defaults 503
Setting Up the Quantification Method for Multiple Input Files 503
Quantification Method Editor Dialog Box Parameters 506
Quantification Methods View Parameters 515
Adding a Quantification Method 516
Create Quantification Method Dialog Box Parameters 519
Changing a Quantification Method 520
Removing a Quantification Method 521
Importing a Quantification Method 522
Exporting a Quantification Method 523
Summarizing the Quantification 523
Quantification Summary Page Parameters 526
Displaying Quantification Spectra 527
Quan Spectra Page Parameters 528
Displaying Quantification Channel Values 529
Displaying Quantification Channel Values for Reporter Ion Quantification 529
Displaying Quantification Channel Values for Precursor Ion Quantification 530
Quantification Channel Values Chart Parameters 532

Displaying the Quantification Spectrum Chart 533
Displaying the Quantification Spectrum Chart for Reporter Ion Quantification 533
Displaying the Quantification Spectrum Chart for Precursor Ion Quantification 538
Quantification Spectrum Chart Parameters 544
Using Reporter Ion Isotopic Distribution Values to Correct for Impurities . . . 545
Excluding Peptides from the Protein Quantification Results 546
Excluding Peptides with High Levels of Co-Isolation 547
Classifying Peptides 548
Calculating Peptide Ratios 550
Understanding the Peptide Ratio Distributions Chart 551
Handling Missing and Extreme Values in Calculating Peptide Ratios 554
Calculating Protein Ratios from Peptide Ratios 557
Case 1: Quantification Result Associated with One Spectrum, One Peptide, and One Protein 557
Case 2: Two Quantification Results Associated with Two Spectra, One Peptide, and One Protein 558
Case 3: Quantification Result Associated with Two Spectra, Two Peptides, and One Protein 558
Case 4: Quantification Result Associated with One Spectrum, Two Peptides, and One Protein 559
Case 5: Quantification Result Associated with One Spectrum, One Peptide, Two Proteins 559
Case 6: Quantification Result Associated with One Spectrum, Two Peptides, and Two Proteins 559
Case 7: Quantification Result Associated with Two Spectra, Two Peptides, and Two Proteins 560
Calculating Ratio Count and Variability 560
Replicates 560
Treatments 561
Ratio Count 561
Ratio Variability 561
Calculating and Displaying Protein Ratios for Multiconsensus Reports 563
Calculating Protein Ratios in Multiconsensus Reports Treated as Treatments 565
Calculating Protein Ratios in Multiconsensus Reports Treated as Replicates 566
Mixed Mode 567
Identifying Isotope Patterns in Precursor Ion Quantification 569
Troubleshooting Quantification 571
Troubleshooting Reporter Ion Quantification 571
Troubleshooting Precursor Ion Quantification 574

Chapter 10 Exporting Data 577
Export Options 577
Exporting Spectra 578
Export Spectra Dialog Box Parameters 579
Exporting Spectra Without a Peptide Match 581
Exporting Annotated Spectra for Selected Proteins or Peptide Groups 582
Peptides Table in the Output HTML File 588
Export Annotated Spectra Dialog Box Parameters 594
Fragment Match Options Dialog Box 595
Exporting Search Results in ProtXML Format 596
Export to ProtXML Dialog Box Parameters 596
Exporting Search Results in PepXML Format 597
To PepXML Dialog Box Parameters 598
Copying or Saving a View to an Image 598
Exporting Protein References to a FASTA File 599
Export to FASTA Dialog Box Parameters 600
Exporting Exclusion Mass Lists to Xcalibur 600
Export Exclusion/Inclusion List Dialog Box Parameters 603
Exporting Search Results to Excel 604
Export to Excel Workbook Dialog Box Parameters 609
Exporting Search Results to a Tab-Delimited TXT File 609
To CSV Dialog Box Parameters 612

Chapter 11 Reference 615
File Menu 616
Search Report Menu 619
Quantification Menu 622
Processing Menu 624
Workflow Editor Menu 625
Administration Menu 626
Tools Menu 628
Window Menu 629
Help Menu 629
Shortcut Menus 630
Proteins Page Shortcut Menu 630
Peptides Page Shortcut Menu 632
Search Input Page Shortcut Menu 635
Graph Shortcut Menu 636
Workflow Editor Shortcut Menu 637

Workflow Editor Nodes 637
Data Input Nodes 638
Spectrum & Feature Retrieval Nodes 639
Spectrum Processing Nodes 647
Spectrum Filters Nodes 659
Peptide/Protein ID Nodes 666
Peptide/Protein Validation Nodes 680
PTM Analysis Nodes 691
Quantification Nodes 698
Data Export Nodes 702
Annotation Nodes 703

Appendix A FASTA Reference 705
FASTA Databases 705
NCBI 705
MSIPI 706
IPI 706
UniRef100 707
SwissProt and TrEMBL 707
MSDB 708
Custom Database Support 708
Custom Parsing Rule A 708
Custom Parsing Rule B 709
Custom Parsing Rule C 709

Appendix B Chemistry References 711
Amino Acid Mass Values 711
Enzyme Cleavage Properties 712
Fragment Ions 713

Appendix C Migrating from BioWorks to the Proteome Discoverer Application 715
Configuring the SRF Importer 715
Installing the Software for the Migration 716
Preparing the Computer 717
Installing BioWorks Software on a Reserved Data System Computer 717
Installing the Proteome Discoverer Application 718
Adding FASTA Databases 718
Installing a Workflow Template 718

Converting SRF, DTA, OUT, and Raw Data 720
Moving BioWorks 3.3.1 SRF Files to the Proteome Discoverer Application. 721
Moving BioWorks 3.3 or Earlier SRF Files to the Proteome Discoverer Application 726
Importing BioWorks DTA, OUT, and FASTA Files into the Proteome Discoverer Application 730
Importing Raw Data Files into the Proteome Discoverer Application 735
Possible Causes of Failed Migrations 737
Known Issues 740

Index 743

Preface

This guide describes how to use the Proteome Discoverer™ 1.3 application for peptide and protein mass spectrometry analyses.

Contents

- Related Documentation
- System Requirements
- Special Notices
- Contacting Us

To provide us with comments about this document, click the link below. Thank you in advance for your help.



Related Documentation

The Proteome Discoverer application includes Help and these manuals as PDF files:

- *Proteome Discoverer User Guide*
- *Proteome Discoverer Installation Guide*

❖ **To view product manuals**

- *Proteome Discoverer User Guide*: Go to **Start > Programs > Thermo Proteome Discoverer 1.3 > Proteome Discoverer 1.3 User Guide**.
- *Proteome Discoverer Installation Guide*: Go to **Start > Programs > Thermo Proteome Discoverer 1.3 > Proteome Discoverer 1.3 Installation Guide**.

❖ **To open Help**

- From the main Proteome Discoverer window, choose **Help > Help Contents**.
- If available for a specific window or view, click **Help** or press F1 for information about setting parameters.

For more information, visit www.thermo.com. You can find application notes at www.thermo.com/appnotes.

System Requirements

The Proteome Discoverer application requires a license. In addition, your system must meet these minimum requirements.

System	Requirements
Hardware	• 2 GHz processor with 2 GB RAM • DVD/R-ROM drive • Video card and monitor capable of 1280 × 1024 resolution (XGA) • Screen resolution of 96 dpi • 75 GB available on the C: drive • NTFS format
Software	• Microsoft Windows™ XP 32/64 Professional (English version) with latest service pack installed • Microsoft Windows 7 32/64 Professional (English version)
Mascot Server	• Mascot Server 2.1 – Mascot servers running version 2.1 should be usable, but retrieving the result files (protein sequences) from the servers can be a lengthy process because you can only retrieve the protein sequences one at a time. – Mascot servers running version 2.1 should have all available updates, patches, or both from Matrix Science installed. In particular, you must install a patch that enables MIME format for the result files; otherwise, the Proteome Discoverer application cannot receive the search results from the Mascot server. • Mascot Server 2.2: Proteome Discoverer 1.3 does not support error-tolerant searches. • Mascot Server 2.3: Proteome Discoverer 1.3 does not support error-tolerant searches, Percolator-based scoring, and searches against multiple-sequence databases.

Note Ensure that port 28199 is not blocked by firewalls.

Note Ensure that the Windows operating system first has the latest Microsoft .NET Framework and Windows updates installed before installing the Proteome Discoverer application.

Special Notices

Make sure you follow the precautionary statements presented in this guide. Special notices appear in boxes.

IMPORTANT Highlights information necessary to prevent damage to software, loss of data, or invalid test results; or might contain information that is critical for optimal performance of the system.

Note Highlights information of general interest.

Tip Highlights helpful information that can make a task easier.

Contacting Us

There are several ways to contact Thermo Fisher Scientific for the information you need.

❖ **To contact Technical Support**

Phone	800-532-4752
Fax	561-688-8736
E-mail	us.techsupport.analyze@thermofisher.com
Knowledge base	www.thermokb.com

Find software updates and utilities to download at mssupport.thermo.com.

❖ **To contact Customer Service for ordering information**

Phone	800-532-4752
Fax	561-688-8731
E-mail	us.customer-support.analyze@thermofisher.com
Web site	www.thermo.com/ms

❖ **To get local contact information for sales or service**

Go to www.thermoscientific.com/wps/portal/ts/contactus.

❖ To copy manuals from the Internet

Go to mssupport.thermo.com, agree to the Terms and Conditions, and then click **Customer Manuals** in the left margin of the window.

❖ To suggest changes to documentation or to Help

- Fill out a reader survey online at www.surveymonkey.com/s/PQM6P62.
- Send an e-mail message to the Technical Publications Editor at techpubs-lcms@thermofisher.com.

Introduction

This chapter introduces you to the Proteome Discoverer application and describes its features and functionality.

Contents

- Features
- Workflow
- Inputs and Outputs
- Limitations
- New Features in This Release

Features

The Proteome Discoverer application identifies proteins from the mass spectra of digested fragments. It compares the raw data from mass spectrometry to the information from the selected FASTA database. You can use this application to analyze spectral data from all Thermo Scientific and other mass spectrometers. Specifically, the Proteome Discoverer application does the following:

- Works with peak-finding search engines such as SEQUEST™, Mascot™, and ZCore™ to process all data types collected from low- and high-mass-accuracy mass spectrometry (MS) instruments. The peak-finding algorithm searches the raw mass spectrometry data and generates a peak list and relative abundances. The peaks represent the fragments of peptides for a given mass and charge.
- Produces complementary data from a variety of dissociation methods and data-dependent stages of tandem mass spectrometry.
- Combines, filters, and annotates results from several database search engines and from multiple analysis iterations. The search engines correlate the uninterrupted tandem mass spectra of peptides with databases, such as FASTA. See "Using FASTA Databases" on page 159.

The Proteome Discoverer application includes the following features:

- Support for multiple search engines:
  - SEQUEST
  - Mascot
  - ZCore
  - SEQUEST, Mascot, and ZCore are available as wizards or as nodes in the Workflow Editor. "Search Engines" on page 3 describes these search engines.
- The Workflow Editor for searching with multiple algorithms and merging results from multiple dissociation techniques. See "Starting a New Search by Using the Workflow Editor" on page 89.
- Support for both precursor ion quantification (for example, SILAC), reporter ion quantification (for example, iTRAQ™ and Tandem Mass Tag™ [TMT]), and peak area calculation quantification. For details, see "Performing Precursor Ion Quantification" on page 469, "Performing Reporter Ion Quantification" on page 476, and "Performing Peak Area Calculation Quantification" on page 483, respectively.
- Access to GO database information from ProteinCenter and Protein Family (Pfam) database information from the Wellcome Trust Sanger Institute to annotate the proteins in your results report (Magellan storage file, or MSF). ProteinCenter is a Web-based application that you can use to download biologically enriched annotation information for a single protein, such as molecular functions, cellular components, and biological processes, from the GO database. For information, see "Protein Annotation" on page 433. You can also upload search results directly from the Proteome Discoverer application to ProteinCenter.
- Proteome Discoverer Daemon, which can perform multiple searches on multiple raw files at any given time. You can use it to perform searches on multiple raw files taken from multiple samples or replicates from the same sample. See "Using the Proteome Discoverer Daemon Utility" on page 121.
- A number of graphical views that contain detailed information about the selected peptides and proteins. You can display more than one view to perform a comparative analysis of your selected peptide or proteins. See "Interpreting Search Results" on page 213.
- The presentation of database search results available from multiple raw files in a single protein or peptide report. See "Creating a Report from Multiple Results" on page 351.
- Support for FASTA databases and indexes. See "Using FASTA Databases" on page 159.
- The ability to import protein and peptide reports in standard spectrum data formats, such as MZDATA, MZXML, MZML, and MGF. See "Importing Raw Data Files in Other Formats into a Workflow" on page 114.

- The ability to export protein and peptide reports in standard spectrum data formats, such as MZDATA, DTA, MZML, and MGF. You can also export search results to XML and tab-delimited TXT files. In addition, you can export annotated spectra for selected peptides into a ZIP file that includes an HTML page with peptide information and links to spectrum images. "Exporting Data" on page 577 describes how to export your data to these and other formats.
- The ability to merge filtered or unfiltered search results. See "Creating a Report from Multiple Results" on page 351.
- A number of protein and peptide filtering and grouping options to help you sort and filter your data. For information on Proteome Discoverer's filtering capabilities, see "Filtering Data" on page 359. For information on grouping, see "Grouping Proteins" on page 401 and "Grouping Peptides" on page 413.

Search Engines

The Proteome Discoverer application includes the SEQUEST, Mascot, and ZCore search engines; each produces complementary data. ZCore and SEQUEST search engines distributed by Thermo Scientific. Mascot is a protein identification search engine created by Matrix Science.

Mascot uses mass spectrometry data to identify proteins from primary sequence databases. ZCore is specifically developed and optimized to evaluate both high-mass-accuracy and low-mass-accuracy ETD data. SEQUEST can analyze different data types.

- Electron-transfer dissociation (ETD)
- Electron-capture dissociation (ECD)
- Collision-induced dissociation (CID)
- High-energy collision-induced dissociation (HCD)
- Pulsed collision-induced dissociation (PQD)
- ETD and ECD generate primarily c and z fragment ions with preferences for precursor ion charge states of +3 or higher.
- CID and HCD generates primarily b and y fragment ions with preferences for precursor ion charge states of +3 or lower.
- PQD and HCD do not exhibit a low-mass cutoff and are good for reporter-ion experiments.

Frequently, peptides identified by CID, PQD, or HCD are not observed with ETD or ECD, and vice versa, so that combining results from, for example, CID and ETD can enhance sequence coverage. Many times CID and ETD identify the same peptides, often with different precursor ion charge states. Combining ETD and CID results improves confidence in identifications.

SEQUEST Search Engine

SEQUEST is specifically developed and optimized to evaluate both high-mass-accuracy and low-mass-accuracy ETD, ECD, CID, HCD, and PQD data. You can use SEQUEST combined with automated LC-MS/MS and intelligent data acquisition tools to ensure the routine identification of low-abundance proteins in complex mixtures.

The Proteome Discoverer application extracts relevant MS/MS spectra from the raw file and determines the precursor charge state and the quality of the fragmentation spectrum.

The SEQUEST search algorithm correlates experimental MS/MS spectra through comparisons to theoretical in-silico peptide candidates derived from protein databases. The proprietary cross-correlation identification algorithm at the core of SEQUEST uses a sophisticated scoring system to help assess results. SEQUEST looks for characteristic spectral patterns and then critically evaluates the equivalence of experimental and theoretical MS/MS spectra. The identification algorithm extracts information and correctly identifies proteins even when protein sample sizes are limited and the signal-to-noise ratio of spectra is low.

You can extract specific information from your results through the interactive data summary screens. With a click, you can examine a fully annotated MS/MS spectrum, or view the percent peptide coverage of an identified protein.

SEQUEST provides excellent search results on data acquired with Thermo Scientific ion trap mass spectrometers. Using accurate mass windows decreases the search time, increases the accuracy of the result, and decreases the false positive rate.

The Proteome Discoverer probability-based scoring system rates the relevance of the best matches found by the SEQUEST algorithm. With this probability-based scoring, the application can independently rank the peptides and proteins and increase the confidence in protein identification. Additionally, this scoring system minimizes the time needed for data interrogation or results review, increasing the overall throughput of the analysis.

You can also automatically determine false discovery rates by comparing the results of forward and reversed databases, which provides an additional means of increasing confidence in protein identification.

Mascot Search Engine

Mascot uses mass spectrometry data to identify proteins from primary sequence databases.

For more details on Mascot, visit http://www.matrixscience.com.

ZCore Search Engine

ZCore is specifically developed and optimized to evaluate both high-mass-accuracy and nominal-mass-accuracy ETD data. You can use ZCore to identify ETD MS/MS fragmentation spectra of precursor ion charge states between 2 and 7.

The ZCore search algorithm correlates experimental MS/MS spectra through comparisons to theoretical in-silico peptide candidates derived from protein databases, even when sample sizes are limited and the signal-to-noise ratio of the spectra is low. ZCore evaluates the quality of the fragmentation spectrum and scores the spectra.

You can extract specific information from your results through the interactive data summary screens. With a click, you can examine a fully annotated MS/MS spectrum or view the percent peptide coverage of an identified protein.

ZCore provides excellent search results on data acquired with Thermo Scientific ion trap mass spectrometers.

To distinguish a correct peptide match from an incorrect peptide match, the ZCore algorithm uses two calculations, a probability-based scoring of the fragment ion distribution and the total fraction of covered fragment ion intensity. The score is reported as a probability (using a -log10 value system) and as an expectation value.

Wizards and Workflow Editor

You can use the Proteome Discoverer application's search wizards or its Workflow Editor to conduct data analysis searches of your spectra.

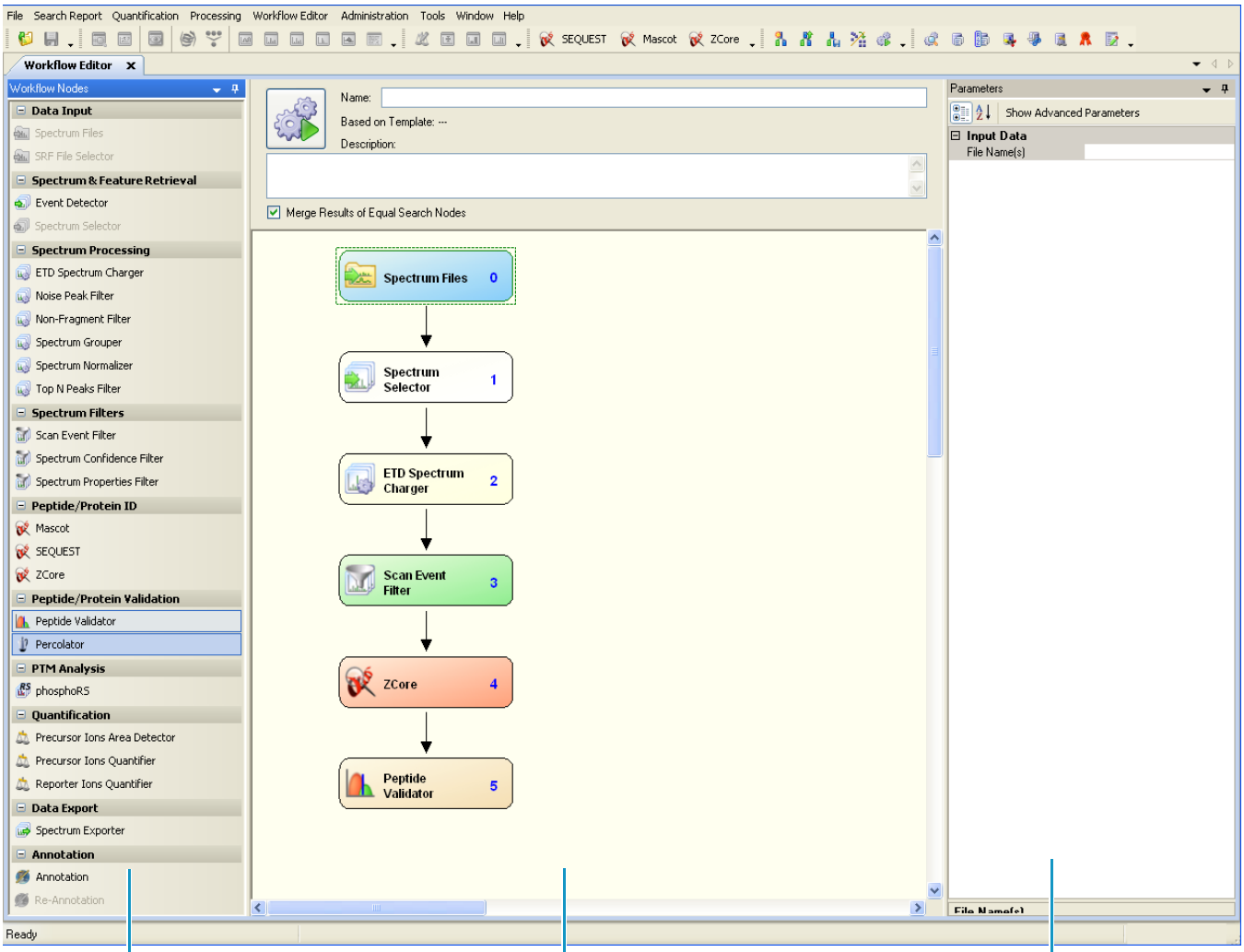
The search wizards are predefined to enable you to quickly set your search parameters and obtain results. The Proteome Discoverer application includes a wizard for each of the supplied search engines: ZCore, SEQUEST, and Mascot.

For information about how to use the wizards, see "Starting a New Search by Using the Search Wizards" on page 63.

The Workflow Editor provides greater flexibility in creating custom search results. Use the three-pane display to create a custom workflow. The Workflow Nodes pane of the application's interface contains seven categories of workflow choices. A typical workflow uses three or more options from these categories, as shown in Figure 1. To start a new workflow, begin with a node from the Data Input category. For more information, see "Starting a New Search by Using the Workflow Editor" on page 89.

When you activate any node from the Workflow Nodes pane, the parameters appear in the Parameters pane.

Figure 1. Workflow Editor workspace



Quantification

The Proteome Discoverer application offers both isotopically labeled precursor ion quantification and isobarically labeled reporter ion quantification methods, which you can also edit.

SILAC is an isotopically labeled quantification method that uses in-vivo metabolic labeling to detect differences in the abundance of proteins in multiple samples. SILAC uses the Precursor Ions Quantifier node in the Workflow Editor.

iTRAQ and TMT are very similar isobarically labeled quantification methods that use external reagents, or tags, to chemically label proteins and peptides to detect differences in abundances. TMT quantification offers default 2plex and 6plex quantification methods, and iTRAQ offers 4plex and 8plex quantification methods. You can use these methods to create your own quantification templates.

iTRAQ and TMT use the Reporter Ions quantifier node in the Workflow Editor.

For detailed information about isobarically and isotopically labeled quantification, see "Performing Reporter Ion Quantification" on page 476 and "Performing Precursor Ion Quantification" on page 469.

The Proteome Discoverer application also offers peak area calculation quantification, which you can use to determine the area for any quantified peptide. This type of quantification uses the Precursor Ions Area Detector node. For more information about peak area calculation quantification, see "Performing Peak Area Calculation Quantification" on page 483.

The Qual Browser Application

With the Qual Browser application, you can view the entire ion chromatogram and browse individual precursor and MS^n data. You can filter the results in a variety of ways, for example, to produce a selected ion chromatogram. When you select a peptide and choose Tools > Open QualBrowser, the Proteome Discoverer application passes the currently active raw file for Qual Browser operations. For more information about the Qual Browser application, see "Using the Qual Browser Application" on page 205.

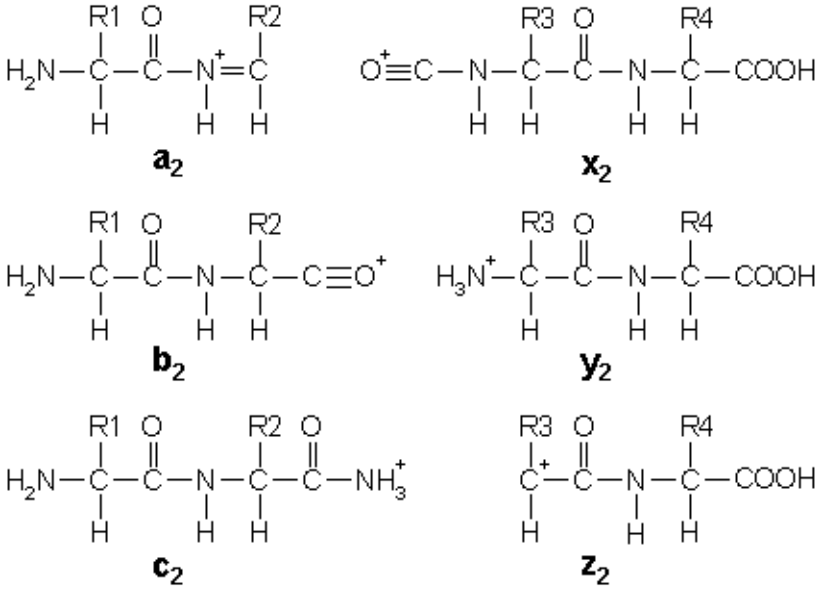
Peptides and Fragment Ions

The types of fragment ions observed in an MS/MS spectrum depend on several factors, such as the primary sequence, the energy source, and the charge state.

Fragment ions of peptides are produced by a collision-induced dissociation (CID) process in which a peptide ion is fragmented in a collision cell. Low-energy CID spectra are generated by MS/MS and ESI, and are sequence-specific. The fragment ion spectra contain peaks of the fragment ions formed by cleavage of the peptide bond and are used to determine the amino acid sequence. A fragment must have at least one charge for it to be detected. If this charge is retained on the N terminal fragment, the ion is classed as a, b, or c. If the charge is retained on the C terminal fragment, the ion type is x, y, or z. A subscript indicates the number of residues in the fragment.

In addition to the proton carrying the charge, c ions and y ions abstract an additional proton from the precursor peptide, as shown in Figure 2.

Figure 2. Structures of six singly charged sequence ions



Fragmentation Methods

The Proteome Discoverer application supports the following fragmentation types:

- CID – Uses the collision-induced dissociation (CID) method of fragmentation, where molecular ions are accelerated to high kinetic energy in the vacuum of a mass spectrometer and then allowed to collide with neutral gas molecules such as helium, nitrogen, or argon. The collision breaks the bonds and fragments the molecular ions into smaller pieces.
- ECD – Uses the electron capture dissociation (ECD) method of fragmentation, where multiply protonated molecules are introduced to low-energy free electrons. Capture of the electrons releases electric potential energy and reduces the charge state of the ions by producing odd-electron ions, which easily fragment.
- HCD – Uses the high-energy collision-induced dissociation (HCD) method of fragmentation, where the projectile ion has laboratory-frame translation energy higher than 1 keV. HCD produces a highly abundant series of reporter ions for TMT and iTRAQ quantification.
- ETD – Uses the electron transfer dissociation (ETD) method of fragmentation, where singly charged reagent anions transfer an electron to multiply protonated peptides within an ion trap mass analyzer to induce fragmentation. ETD cleaves along the peptide backbone while side chains and modifications such as phosphorylation are left intact. This method is used to fragment peptides and proteins.

- IRMPD – With the infrared multi-photon dissociation (IRMPD) method of fragmentation, an infrared laser is directed at the ions in the vacuum of the mass spectrometer. The target ions absorb multiple infrared photons until they reach more energetic states and begin to break bonds, resulting in fragmentation.
- PQD – Uses the pulsed Q collision-induced dissociation (PQD) method of fragmentation, where precursor ions are activated at a high value, a parameter that determines the stability of an ion's trajectory in an ion trap mass analyzer. Then, a time delay occurs to allow the precursor to fragment, and then a rapid pulse is applied to a low value where all fragment ions are trapped. The product ions can then be scanned out of the ion trap and detected. PQD fragmentation produces precise, reproducible fragmentation and has been used for iTRAQ peptide quantification on the LTQ™ mass spectrometer using both electrospray and MALDI source ionization.

MudPIT Experiments

Multidimensional Protein Identification Technology (MudPIT) experiments investigate complex proteomes by applying multidimensional chromatography to the samples before acquisition in the mass spectrometer. Typically, this process results in several dozen or even a few hundred fractions that are separately analyzed by LC-MS, resulting in one raw file per sample fraction. Analyzing gel slices or performing in-depth follow-up acquisitions also results in multiple fractions. Because all these fractions belong to the same sample, the Proteome Discoverer application can process all raw files from these fractions as one contiguous input file and generate a single result file. For detailed information about processing MudPIT samples, see "Using the Proteome Discoverer Daemon Utility" on page 121.

Workflow

Through settings that you specify in the Proteome Discoverer application, you can search, filter, and sort raw files with the SEQUEST, Mascot, and ZCore algorithms. In addition to creating reports from the analyzed data, the application extracts relevant MS/MS spectra from the raw file and determines the precursor charge state. Filters in the application remove false positives and other irrelevant information with a variety of user-specified methods.

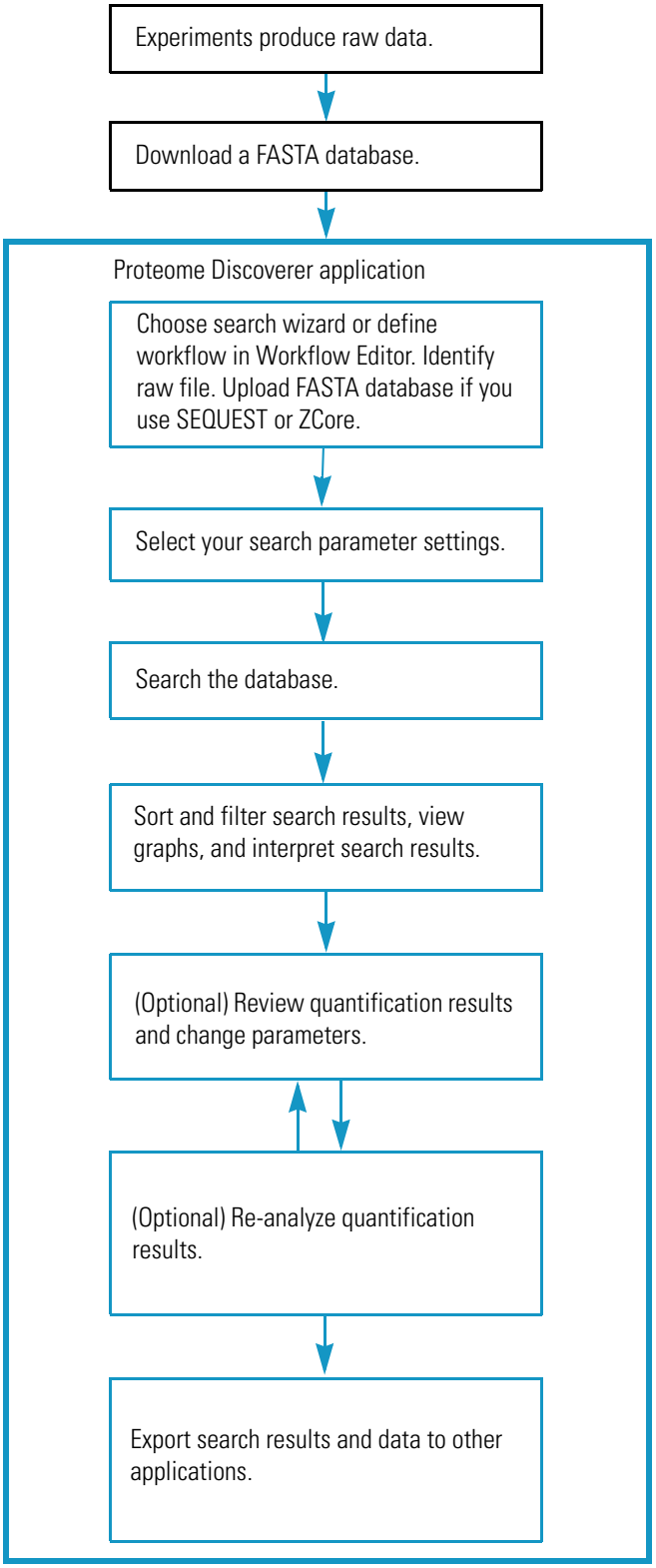
Note You can filter data according to false discovery rates that you define through the use of decoy databases that you specify in the workflow.

Using the standard Proteome Discoverer workflow involves the following steps when you process, analyze, and interpret mass spectrometry data. These steps are shown graphically in Figure 3.

1. Upload a FASTA database, if necessary, if you are using SEQUEST or ZCore.
2. Choose a search wizard or create a workflow in Workflow Editor. Identify the raw file.
3. Select parameter settings in the search wizard or the nodes of the Workflow Editor.

4. Begin a search of the raw data. The Proteome Discoverer application initiates a search against a FASTA database.
5. Sort and filter the search report, generate graphs and views, and interpret the search results.
6. (Optional) Review the quantification results and change parameters.
7. Reanalyze the quantification results.

Figure 3. The Proteome Discoverer workflow



Inputs and Outputs

The Proteome Discoverer application can accept several different file formats as input and can export data in several formats.

FASTA Databases

The Proteome Discoverer application includes FASTA databases, including multiple example FASTA databases and example raw files. Use these files when exploring and learning how to use the application. For a detailed description of the different types of FASTA databases and their purpose, see "Using FASTA Databases" on page 159.

Inputs

The Proteome Discoverer application accepts the following file types as input:

- Xcalibur raw files contain raw data collected from a mass spectrometer.
- Mascot Generic Format (MGF) files are mass spectral files produced during Mascot analysis. They contain a list of precursor ions, their fragments, and the masses of the fragments.
- Extensible Markup Language (XML) files contain workflow templates.
- MZXML files are standard 2.*x* mass spectrometer data format files, developed at the Seattle Proteome Center at the Institute for Systems Biology (ISB), that contain a list of precursor ions, their fragments, and the masses of the fragment.
- MZDATA files are common data format files developed by the Human Proteome Organization (HUPO) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line.
- MZML files are a combination of .mzData and .mzXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the Seattle Proteome Center at the Institute for Systems Biology (ISB). The Proteome Discoverer application supports version 1.1.0 of the MZML format.
- Magellan storage (MSF) files contain the results of the searches conducted by the search wizards or the Workflow Editor.
- Unified search results (SRF) files from the Thermo Scientific BioWorks™ application. See "Migrating from BioWorks to the Proteome Discoverer Application" on page 715 for information about converting these files for use in the application.

Outputs

The Proteome Discoverer application creates the following file types as output:

- DTA Archive (DTA) files are files containing MS^n data for single or grouped scans.
- Mascot Generic Format (MGF) files are mass spectral files produced during Mascot analysis. They contain a list of precursor ions, their fragments, and the masses of the fragments.
- MZDATA files are common data format files developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line.
- Magellan storage (MSF) files contain the results of the searches conducted by the search wizards or the Workflow Editor.
- Extensible Markup Language (XML) files contain workflow templates.
- MZXML files are standard 2.*x* mass spectrometer data format files developed at the Seattle Proteome Center at the Institute for Systems Biology (ISB) that contain a list of precursor ions, their fragments, and the masses of the fragment.
- MZML files are a combination of MZDATA and MZXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the Seattle Proteome Center at the Institute for Systems Biology (ISB). The Proteome Discoverer application supports version 1.1.0 of the MZML format.
- ProtXML files contain protein identifications from MS/MS-derived peptide sequence data. They are created by the File > Export > To ProtXML command.
- PepXML files contain peptides that are included in the results of searches performed by the SEQUEST and Mascot search engines. They are in PepXML format version 1.14, which is an open data format developed by SPC/Institute for Systems Biology for storing, exchanging, and processing peptide sequence assignments from MS/MS scans. PepXML files are created by the File > Export > pepXML command.

  The Proteome Discoverer application supports version 1.14.
- Tab-delimited TXT files are in a simple text format that stores tabular data and is widely used to exchange data between different computer programs.

Limitations

This release of the Proteome Discoverer application has the following limitations:

- The spectra count is not directly available in the application results report. However, the number of identified peptides is displayed for each protein. This number should be similar to the spectra count for that protein.
- The Proteome Discoverer application supports peptide quantification methods that use reporter ions. Examples of these methods are TMT and iTRAQ. The application also supports peptide quantification methods that measure precursor ion abundances. Examples of these methods are SILAC, ICPL, ^{18}O, ^{15}N, and label-free methods.

New Features in This Release

Proteome Discoverer version 1.3 adds the following new features.

Protein Annotation

You can annotate results files with information from ProteinCenter, a Web-based application that you can use to download biologically enriched annotation information for a single protein, such as molecular functions, cellular components, and biological processes, from the GO and Pfam databases. In addition, you can upload protein results directly from the Proteome Discoverer application to ProteinCenter. For details on this new protein annotation functionality, see "Protein Annotation" on page 433.

Data Display

This release enhances the Proteome Discoverer application's data display capabilities. In previous versions, you could display X, Y plots of basic PSM-level data, but now you can plot data from different result categories: protein groups, proteins, peptide groups, peptide spectrum matches (PSMs), and search input. You can also plot quantification data like quantification ratios and variabilities. In addition to scatter plots, you can now display your data in Venn diagrams and pie charts, and you can customize all these charts. For details, see "Interpreting Your Results with the Report Item Distribution Chart" on page 257.

You can use the new Protein Group Members view to show which proteins are included in a protein group. See "Interpreting Your Results with the Protein Group Members View" on page 285.

New columns have been added to the Proteins, Peptides, and Search Input pages. See "Proteins Page Parameters" on page 241, "Peptides Page Parameters" on page 291, and "Search Input Page Parameters" on page 347, respectively, for information on the columns displayed on these pages.

PTM Analysis

The Protein Identification Details view has been enhanced to display biologically relevant post-translational modifications (PTMs). For each protein, this view can display information retrieved from ProteinCenter, which accesses publicly available databases such as UniProt. The site probabilities of the phosphorylations are also available if you use the phospho*RS* node in the workflow that generates the results. For details, see "Interpreting Your Results with the Protein Identification Details View" on page 245 and "phosphoRS Node" on page 691.

The Chemical Modifications view can now display the type, or category, of the modifications. See "Opening the Chemical Modifications View" on page 196 and "Adding Amino Acids" on page 201.

Data Exportation

This release adds a number of options for exporting data:

- You can export search results to a tab-delimited TXT file. The Proteome Discoverer application exports each page (Proteins, Peptides, or Search Input) of the MSF file in a separate file. For more information, see "Exporting Search Results to a Tab-Delimited TXT File" on page 609.
- You can export all or filtered proteins in a search report to an XML file in ProtXML format. The exported data includes the reporter ion quantitation ratios, precursor ion quantitation ratios, or both, if they are available. See "Exporting Search Results in ProtXML Format" on page 596.
- You can export search results from all search engines, including ZCore, in PepXML format. See "Exporting Search Results in PepXML Format" on page 597.
- New options are available when you export annotated spectra. See "Exporting Annotated Spectra for Selected Proteins or Peptide Groups" on page 582.
- You can export spectra from the search results to MZML file format. See "Exporting Data" on page 577.

Protein Grouping

This release improves the algorithm that the Proteome Discoverer application uses to group proteins. See "Grouping Proteins" on page 401 for detailed information on this algorithm and the application's protein-grouping features.

Peptide Grouping

The Peptides page in the MSF file now displays peptides grouped by default. You can still view the peptides that constitute the group in the new Show Peptides Group Members view, which is accessible through a command on the Search Report menu. See "Grouping Peptides" on page 413 and "Interpreting Your Results with the Peptide Group Members View" on page 343.

Peptide Confidence

This release adds two new nodes, Percolator and Peptide Validator, to improve the confidence of peptide identification in the search results. For information on these two nodes, see "Percolator Node" on page 681 and "Peptide Validator Node" on page 680, respectively.

Workflow Editor Nodes

This release adds several new nodes to the Workflow Editor.

- The Percolator node uses semi-supervised machine learning to discriminate correct from incorrect PSMs and calculates accurate statistics such as the q-value (FDR) and posterior error probability to improve the number of confidently identified peptides at a given false discovery rate. For details, see"Percolator Node" on page 681.
- The phospho*RS* node provides a confidence measure for the localization of phosphorylation in peptide sequences. For details, see "phosphoRS Node" on page 691.
- The Top *N* Peaks Filter node improves the number of peptides identified with high confidence by reducing the number of peaks in the searched peak lists. For details, see "Top N Peaks Filter Node" on page 658.
- The Spectrum Confidence Filter node sets up iterative searches. For details, see "Spectrum Confidence Filter Node" on page 664.
- The Peptide Validator node calculates the probability that the search algorithm incorrectly included a peptide in a sample. This node is used only for peptides. For details, see "Peptide Validator Node" on page 680.
- The Annotation node retrieves GO database information from ProteinCenter and installs it in Proteome Discoverer results files. For details, see "Annotation Node" on page 704.
- The Re-Annotation node updates existing annotations or annotates existing MSF files that do not yet include annotations. For details, see "Re-Annotation Node" on page 704.

Fragment Matching

The Proteome Discoverer application only calculates water neutral loss fragments if side chains contain the potential neutral loss groups of aspartic acid, glutamic acid, serine, or threonine. It only calculates ammonia neutral loss fragments if side chains contain the potential neutral loss groups of lysine, asparagine, glutamine, or arginine.

64-Bit Version

The Proteome Discoverer application is now available in both 32-bit and 64-bit versions.

Separate installers are available for the two different platforms. If you are installing the software from the DVD, the XInstall program automatically chooses the Proteome Discoverer version that matches your version of Microsoft Windows. On a 32-bit Windows, it installs the 32-bit version of Proteome Discoverer, and on a 64-bit version of Windows, it installs the 64-bit version. For details, refer to the *Proteome Discoverer Installation Guide*.

Manual Creation of Decoy Databases

The Proteome Discoverer application automatically creates FASTA indexes for a full enzymatic digestion during a SEQUEST or ZCore search, if an adequate FASTA index does not already exist. Now you can manually create FASTA indexes for semi-enzymatic or no-enzyme searches. See "Manually Creating FASTA Indexes" on page 189.

Using the Proteome Discoverer Interface

This chapter describes the Proteome Discoverer user interface and how to customize it.

Contents

- Proteome Discoverer User Interface
- Customizing the User Interface

Proteome Discoverer User Interface

This section describes the following features of the Proteome Discoverer user interface:

- Main Window
- Colors
- Menu Bar
- Shortcut Menus
- Toolbar
- Status Bar
- Job Queue
- Search Results Window Icons
- Dialog Box Icons

Main Window

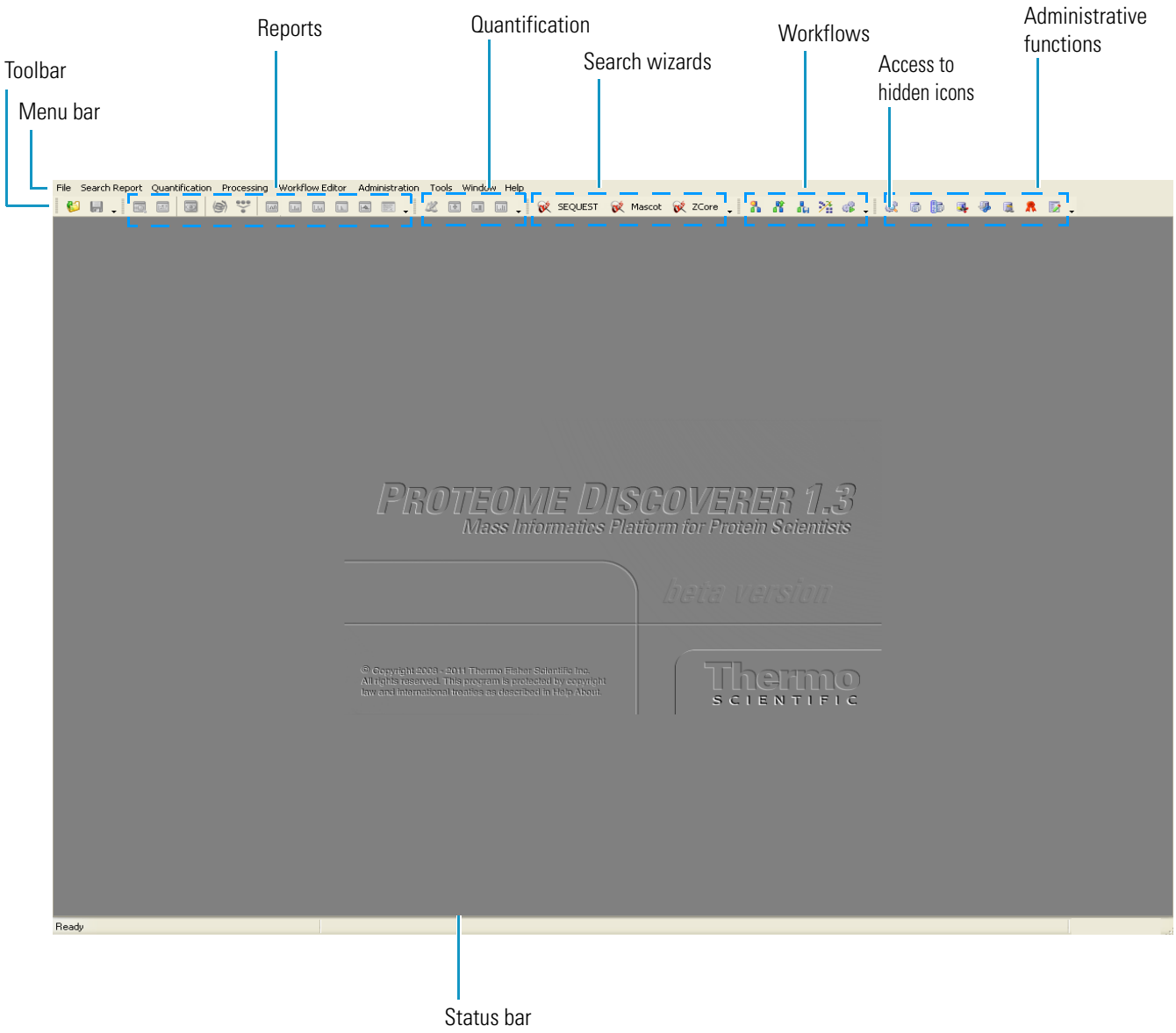
Use the main window of the Proteome Discoverer application to process data, view reports, customize the toolbar, and start various tools.

When you first open the Proteome Discoverer application, the toolbar options and icon names appear in the main window, as shown in Figure 4 (for information on the toolbar icons, see "Toolbar" on page 24). The icons give you quick access to the following.

- Views and graphs
- Search wizards
- Administrative options and features

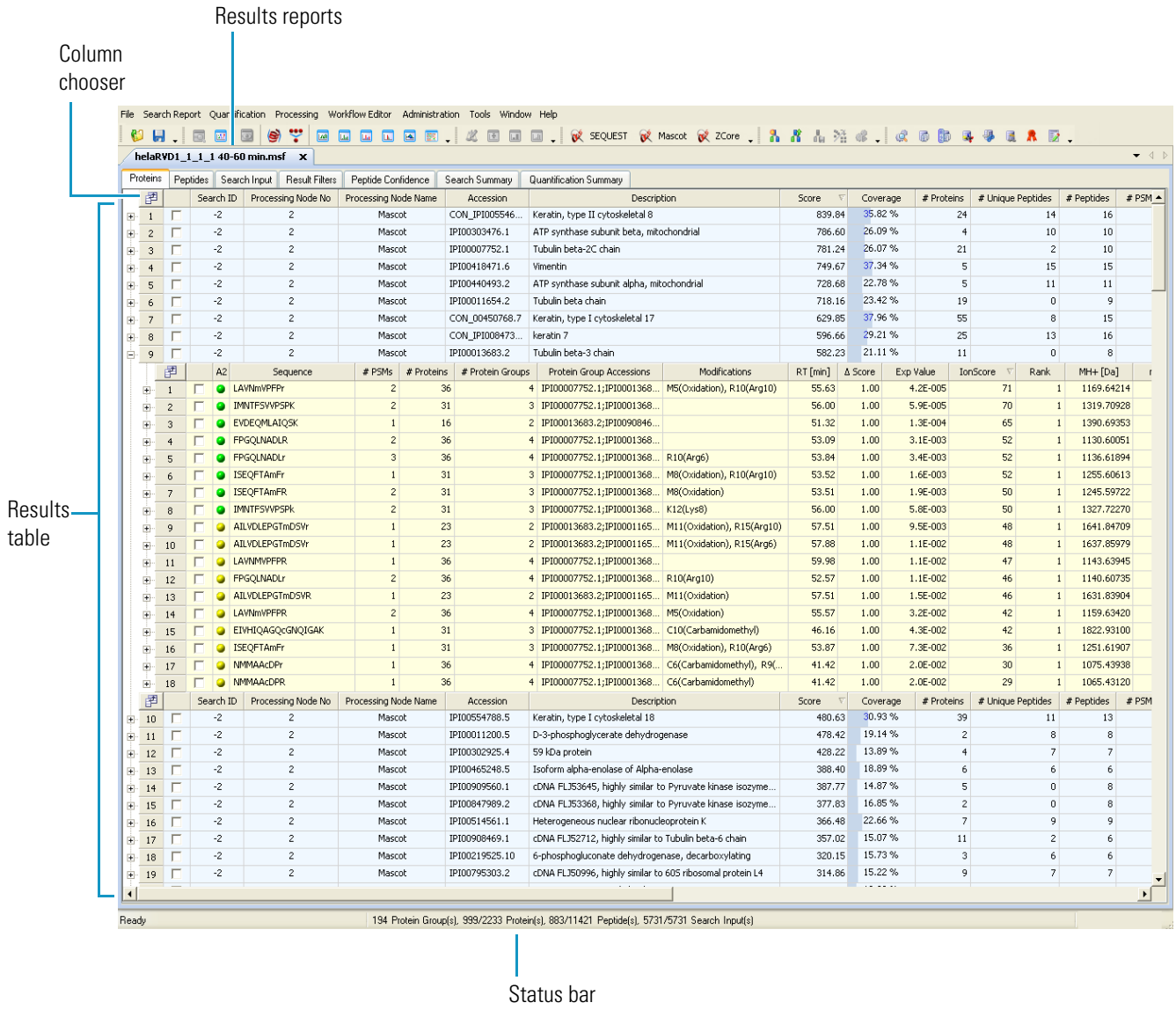
Use the toolbar handles to move a group of icons to a different location on the toolbar. Use the down arrow icons, , to access the Customize menu, which provides options to change the visual appearance of the application.

Figure 4. Proteome Discoverer window



You can open two or more types of reports, in addition to the administration page, at the same time, as shown in Figure 5, but only one page at a time is activated.

Figure 5. Results reports displayed in the Proteome Discoverer window



Colors

The colors on the pages of the Proteome Discoverer application signify the following:

- Blue rows indicate proteins.
- Orange rows indicate ungrouped peptides.
- Green rows are the default color of the Search Input page.
- Yellow rows indicate that peptides have been grouped with the Show Peptide Groups command on the shortcut menu.

- Light gray rows highlight proteins or peptides that have been filtered out by filters on the Result Filters page and then displayed with the Show Filtered Out Rows command on the shortcut menu.
- Darker gray rows highlight proteins or peptides that have been filtered out by row filters and then displayed with the Show Filtered Out Rows command on the shortcut menu.
- Green circles to the left of a peptide sequence indicate a high-confidence level.
- Yellow circles to the left of a peptide sequence indicate a medium-confidence level.
- Red circles to the left of a peptide sequence indicate a low-confidence level.

Cells in the Quan Info column on the Proteins and Peptides pages can have the following background colors:

- Green: Indicates that the peptide is used to calculate the protein ratio.
- Red: Indicates one of the states described for the Quan Info column in Table 45.

Menu Bar

The Proteome Discoverer menu bar at the top of the application interface contains the menus listed in Table 1. For detailed descriptions of each command on these menus, see "Reference" on page 615.

Table 1. Proteome Discoverer menus

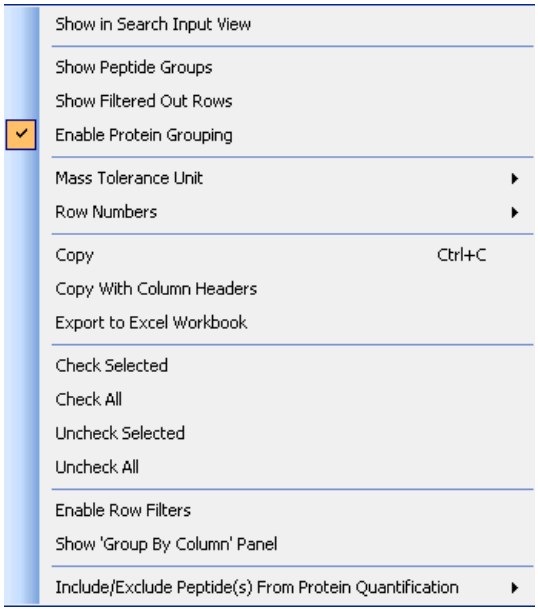
Item	Description
File	Opens, saves, and closes reports; connects to and disconnects from a server; imports results from an SRF file; exports search data; and exits the Proteome Discoverer application.
Search Report	Displays search data.
Quantification	Modifies quantification parameters and interprets results.
Processing	Starts the search wizards that perform searches.
Workflow Editor	Creates, imports, and exports a customized search workflow.
Administration	Performs necessary tasks not directly involved in searching for peptides or proteins, such as displaying the job search queue or downloading FASTA files.
Tools	Opens Qual Browser and a FASTA database management tool.
Window	Closes a window or brings a window forward.
Help	Opens the Help and displays licensing and version information.

Shortcut Menus

The Proteome Discoverer application reports and views have shortcut menus with features that you can use to help analyze data.

- In a report, right-click anywhere in the report to display its shortcut menu. Figure 6 shows the shortcut menu for the Peptides page.

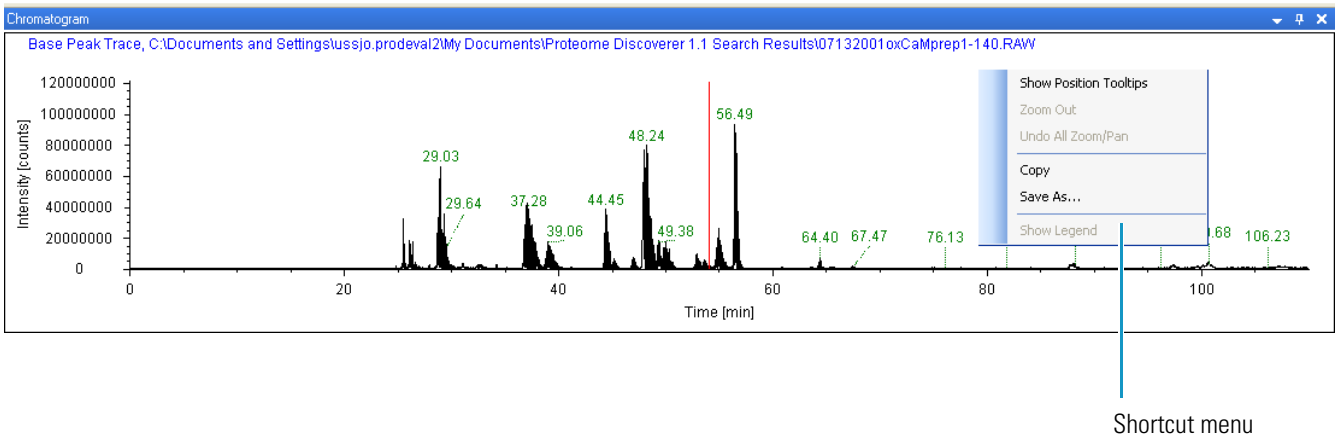
Figure 6. Report shortcut menu



For a complete description of all the shortcut menus available in the Proteome Discoverer application, see "Shortcut Menus" on page 630.

- In the graphs displayed by the commands on the Search Report menu, right-click anywhere to display a shortcut menu containing commands that let you zoom into the chromatogram, copy it, or save the view. Figure 7 shows an example of a graph shortcut menu.

Figure 7. Graph shortcut menu



For a description of the commands on the graph shortcut menu, see "Graph Shortcut Menu" on page 636.

Toolbar

The Proteome Discoverer toolbar, pictured in Figure 8, resides directly below the menu bar on the application interface.

Figure 8. Proteome Discoverer toolbar

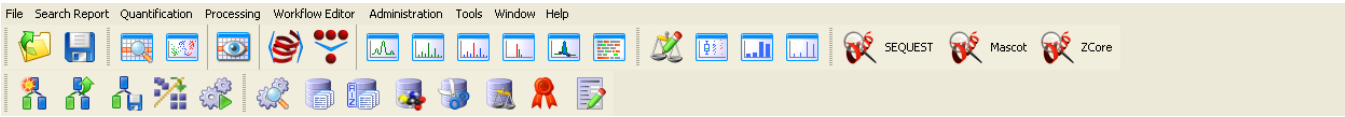


Table 2 describes the function of each toolbar icon.

Table 2. Proteome Discoverer toolbar icons (Sheet 1 of 3)

Toolbar icon	Function	Menu equivalent	Command-line equivalent
	Opens a report.	File > Open Report	CTRL+O
	Saves a modified MSF file, specifically: • The peptide confidence and FDR settings of the Peptide Confidence page • The information in the Annotation column of the Peptides page • The information resulting from the setting that you selected by right-clicking on the peptide rows and choosing the Include/Exclude Peptide(s) from Protein Quantification command (see Table 112 for information on the settings available with this command)	File > Save Report	CTRL+S
	Displays the Protein Identification Details view when you select a protein row or the Peptide Identification Details view when you select a peptide row.	Search Report > Show Protein ID Details or Show Peptide ID Details	CTRL+D
	Displays the Report Item Distribution chart.	Search Report > Show Distribution Chart	CTRL+SHIFT+D

Table 2. Proteome Discoverer toolbar icons (Sheet 2 of 3)

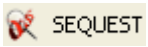
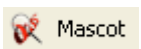
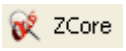
Toolbar icon	Function	Menu equivalent	Command-line equivalent
	Displays the Peptide Consensus view for the selected peptide.	Search Report > Show Peptide Consensus View	
	Displays the Protein Group Members view.	Search Report > Show Protein Group Members view	CTRL+SHIFT+H
	Displays the Peptides Group Members view.	Search Report > Show Peptides Group Members view	
	Displays the Chromatogram view.	Search Report > Show Chromatogram View	CTRL+SHIFT+C
	Displays the Spectrum view.	Search Report > Show Spectrum	CTRL+SHIFT+S
	Displays the Fragment Match Spectrum view.	Search Report > Show Fragment Match Spectrum	CTRL+SHIFT+F
	Displays the Precursor Isotope Pattern view.	Search Report > Show Precursor Isotope Pattern	CTRL+SHIFT+P
	Displays the Extracted Ion Current (XIC) Chromatogram view.	Search Report > Show Extracted Ion Chromatogram	CTRL+SHIFT+T
	Displays the Sequence Comparison view.	Search Report > Show Sequence Comparison	
	Opens the Quantification Method Editor dialog box so that you can edit the quantification method of the current report.	Quantification > Edit Quantification Method	
	Shows the peptide ratios per protein view.	Quantification > Show Peptide Ratios	CTRL+SHIFT+R
	Shows the quantification channel values.	Quantification > Show Quan Channel Values	CTRL+SHIFT+N
	Shows the spectrum used for quantification.	Quantification > Show Quantification Spectrum	CTRL+SHIFT+Q
	Starts the SEQUEST wizard.	Processing > Start SEQUEST Search Wizard	CTRL+Q
	Starts the Mascot wizard.	Processing > Start Mascot Search Wizard	CTRL+M

Table 2. Proteome Discoverer toolbar icons (Sheet 3 of 3)

Toolbar icon	Function	Menu equivalent	Command-line equivalent
	Starts the ZCore wizard.	Processing > Start ZCore Search Wizard	CTRL+SHIFT+Z
	Opens a new workflow in the Workflow Editor.	Workflow Editor > New Workflow	
	Opens an existing workflow in the Workflow Editor.	Workflow Editor > Open From Template	
	Saves a workflow as a template.	Workflow Editor > Save As Template	
	Automatically adjusts and aligns the connecting arrows and nodes.	Workflow Editor > Auto Layout	
	Begins the data analysis search using your selected workflow.	Workflow Editor > Start Workflow	
	Displays the job queue.	Administration > Show Job queue	CTRL+J
	Adds, removes, and modifies FASTA files.	Administration > Maintain FASTA Files	
	Adds, removes, and changes the location of FASTA indexes.	Administration > Maintain FASTA Indexes	
	Allows you to enable or disable available chemical modifications to be used in the search process, create new modifications, or edit existing modifications.	Administration > Maintain Chemical Modifications	
	Adds, removes, and modifies the cleavage reagents.	Administration > Maintain Cleavage Reagents	
	Maintains and edits available quantification methods.	Administration > Maintain Quantification Methods	
	Displays the license status and allows you to add new licenses.	Administration > Manage Licenses	
	Configures workflow node-specific parameters used in searches.	Administration > Configuration	

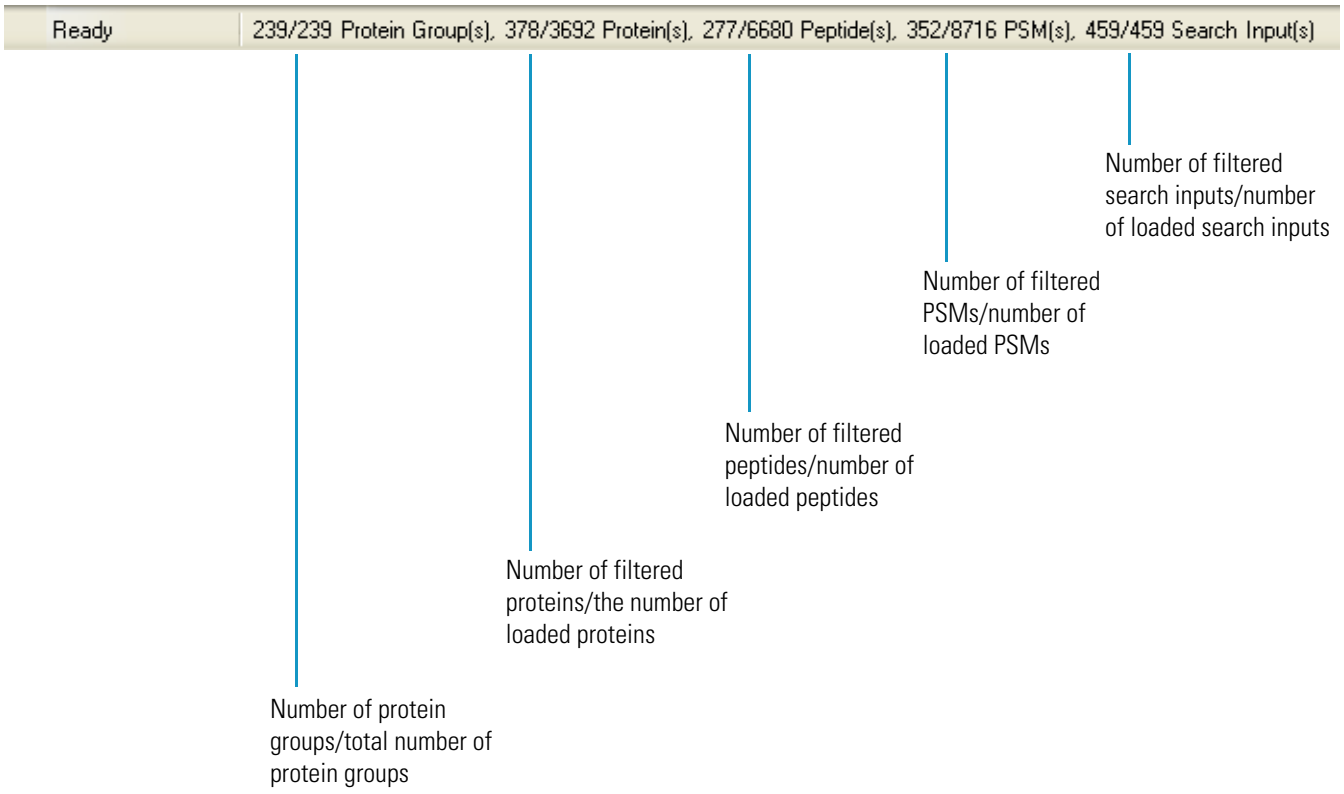
For information about customizing the toolbar, see "Customizing the Toolbar" on page 36 and "Customizing Toolbar Icons, Fonts, and ToolTips Display" on page 41.

Status Bar

The status bar, shown in Figure 9, is located at the bottom of the main window. It displays the following statistics about the loaded report:

- The number of protein groups compared to the total number of protein groups. The difference is the number of protein groups that the application removed to comply with the selection of the Apply Strict Maximum Parsimony Principle option on the Results Filters page. For more information on protein grouping, see "Grouping Proteins" on page 401.
- The number of filtered proteins compared to the number of loaded proteins.
- The number of filtered peptides compared to the number of loaded peptides. The number of peptides in the status bar is always the number of distinct sequences. For more information on the peptides in the status bar, see "Grouping Peptides" on page 413.
- The number of filtered PSMs compared to the number of loaded PSMs.
- The number of filtered search inputs compared to the number of loaded search inputs.

Figure 9. Status bar



Job Queue

You can use the job queue, shown in Figure 10 on page 29, to check on your search status and to control your search progress. It displays the status of any previously run or currently running searches in a Proteome Discoverer application session. You can also use the job queue to open previously run search reports. After you set up and start the search project, you can check the status of your search or remove search results. The job queue displays the search name, the raw file name, the date that the search was submitted, and the progress of the search and its status.

In the job queue, you can do the following:

- Check the status of a search and the results.
- Refresh the job queue status.
- Review result details.
- Pause or resume a search that has already started running.
- Remove or cancel a search.
- Open a report in the job queue.

You can also monitor job execution in the Discoverer Daemon application when you want to perform multiple searches on multiple raw files at any given time. For instructions on this procedure, see "Using the Proteome Discoverer Daemon Utility" on page 121.

To customize the toolbar of the job queue, see "Customizing the Job Queue Toolbar" on page 51.

Displaying the Job Queue

The job queue automatically appears when the Proteome Discoverer application begins processing a workflow, but you can also open it at any time.

❖ To display the job queue

1. Choose **Administration > Show Job Queue**, or click the **Show Job Queue** icon,  .

   The job queue opens on the Administration page of the application interface, as shown in Figure 10.

Figure 10. Job queue showing search with single input file

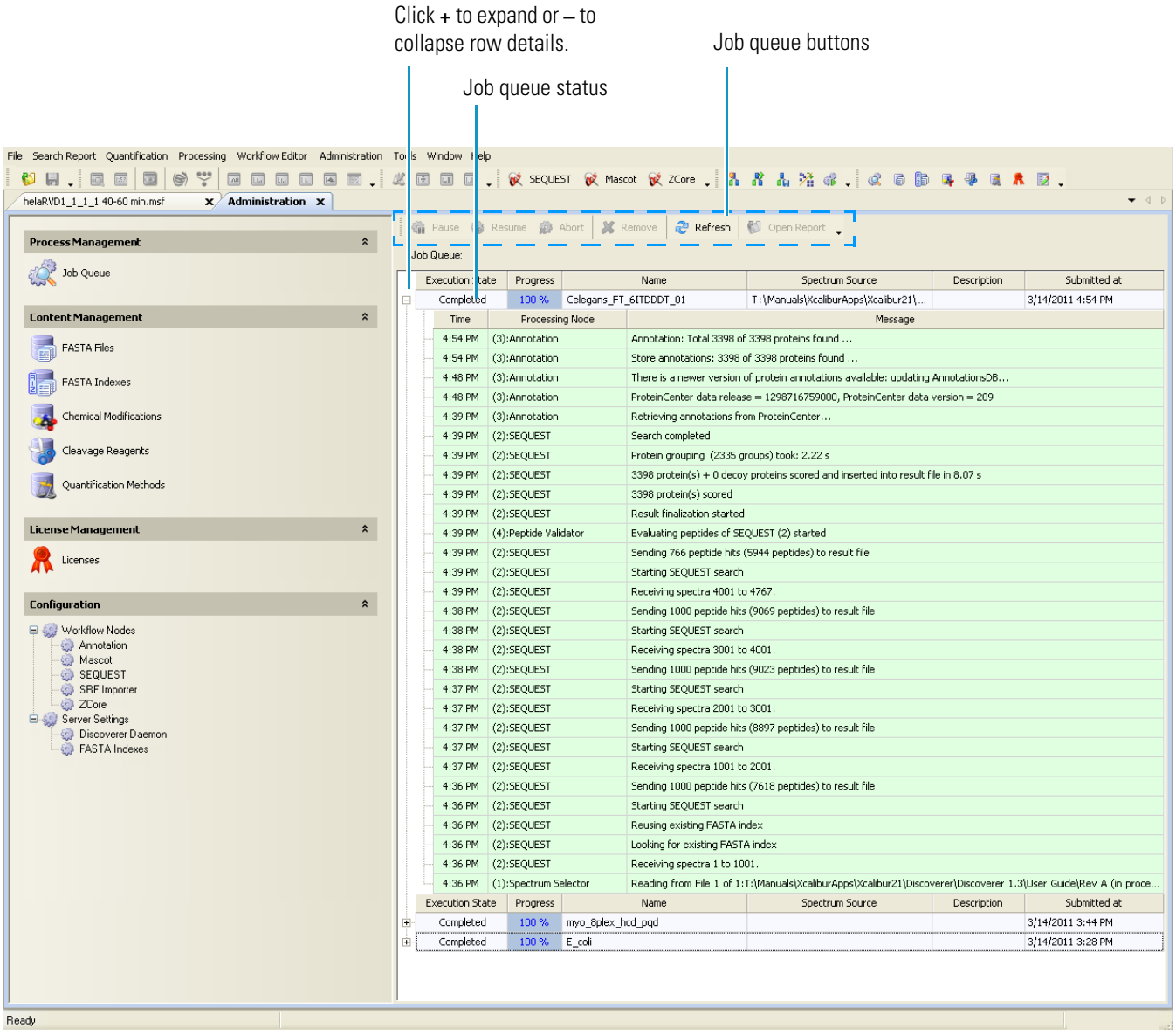
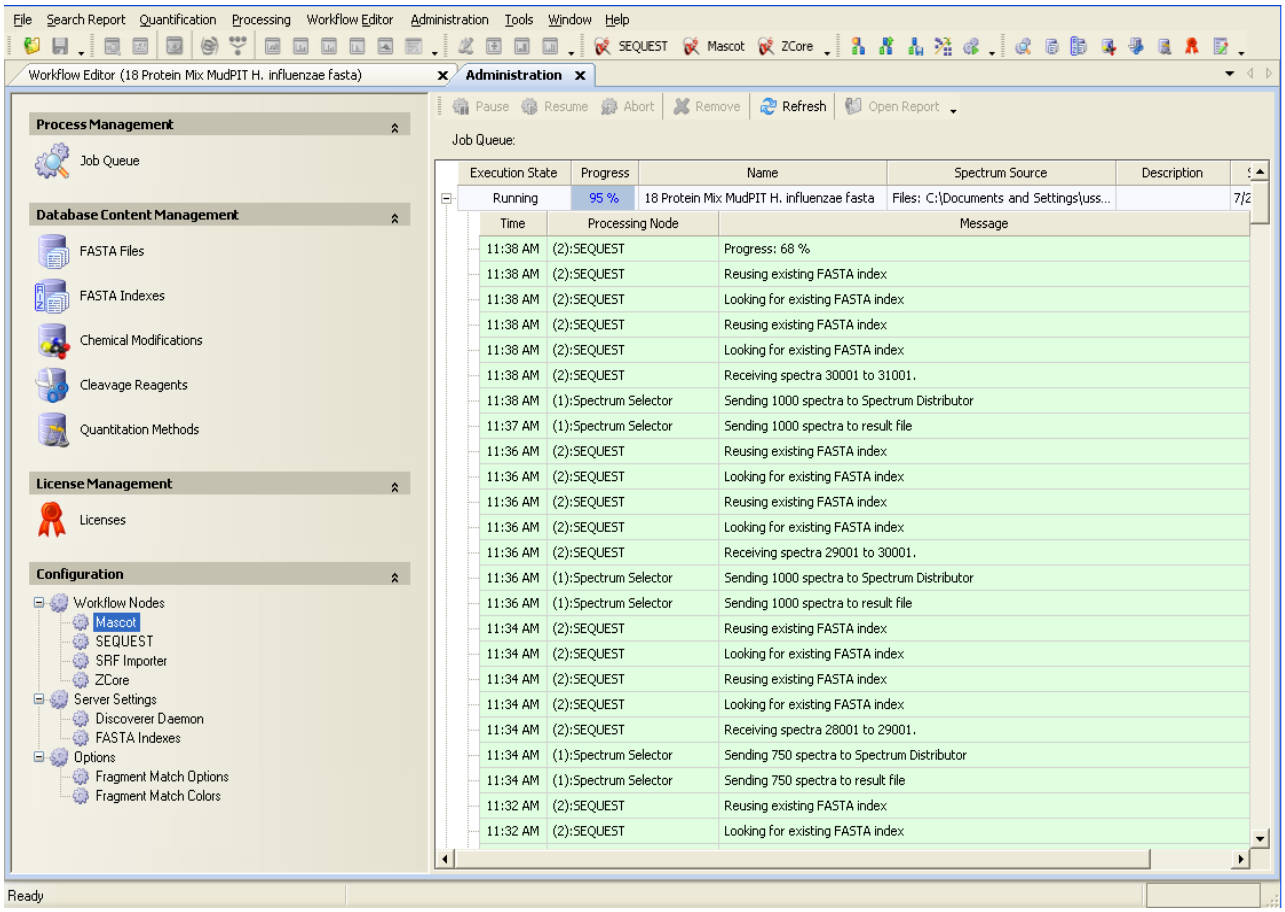


Figure 10 shows a search that has a single input file. Figure 11 shows a search that has multiple input files. In this case, all file names are concatenated and prefixed by "Files:" in the Spectrum Source column.

Figure 11. Job queue showing search with multiple input files



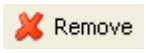
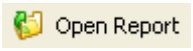
2. To expand and view the details of the search, click + to the left of the search name.

Table 3 describes the job queue buttons available to start or stop a search in progress and job queue status features.

Table 3. Job queue parameters (Sheet 1 of 2)

Parameter	Description
Pause	Temporarily suspends a waiting search until you click Resume. You can only suspend or resume search jobs that are in the Waiting state.
Resume	Places a paused search back into the queue.
Abort	Stops a search.

Table 3. Job queue parameters (Sheet 2 of 2)

Parameter	Description
	Removes a search.
	Refreshes the display of the job queue.
	Opens a completed search report.
Execution State	Displays the status of the search: • Completed: The search is finished. • Not Queued: The search is waiting to be put into the queue because the Proteome Discoverer application is temporarily occupied with another task. • Running: The search is being run now. • Waiting: The search is waiting to run. • Sleeping: The search is in the queue but is suspended. You can place a search in the Sleeping state by selecting a search in the Waiting state and clicking Pause. You can resume the search by selecting the search in the Sleeping state and clicking Resume. • Execution Failed: The search cannot finish.
Progress	Displays the completion of the job as a percentage.
Name	Displays the name of your project from the name that you specified in the search wizard or the Workflow Editor.
Spectrum Source	Displays the path and name of the raw file being used in the search.
Description	Displays the additional descriptive information that you entered when you used the search wizard or the Workflow Editor. This field is blank if you did not add a description.
Submitted at	Displays the date and time that you submitted the search.
Time	Displays the time that the processing finished for each step described in the Message column.
Processing Node	Displays the name of the workflow node performing the step described in the Message column.
Message	Displays the nature of each step in the processing.

Deleting Searches from the Job Queue

You can remove searches that you no longer need from the job queue.

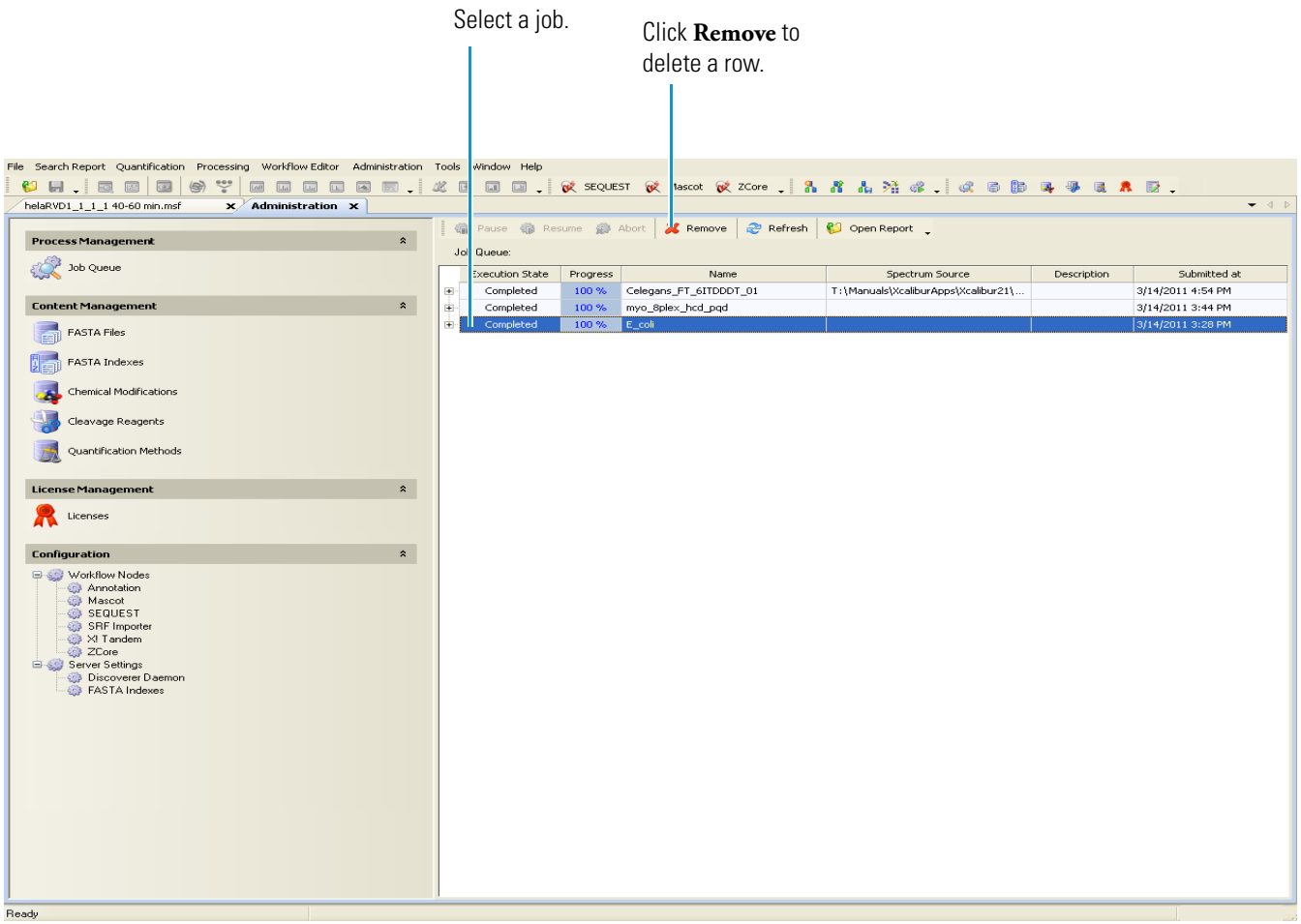
❖ To delete searches from the job queue

Note Jobs are automatically removed from the queue after one week.

1. Select the job that you want to delete, as shown in Figure 12.

   The search is selected.

Figure 12. Removing an item from the job queue



2. Click **Remove** on the job queue toolbar.

   A Delete Jobs confirmation box appears.

3. Click **OK.**

   The Proteome Discoverer application removes the search.

Opening a Completed Report in the Job Queue

You can open a report that you recently worked on in the job queue.

1. In the job queue, click the Completed row that you are interested in viewing.

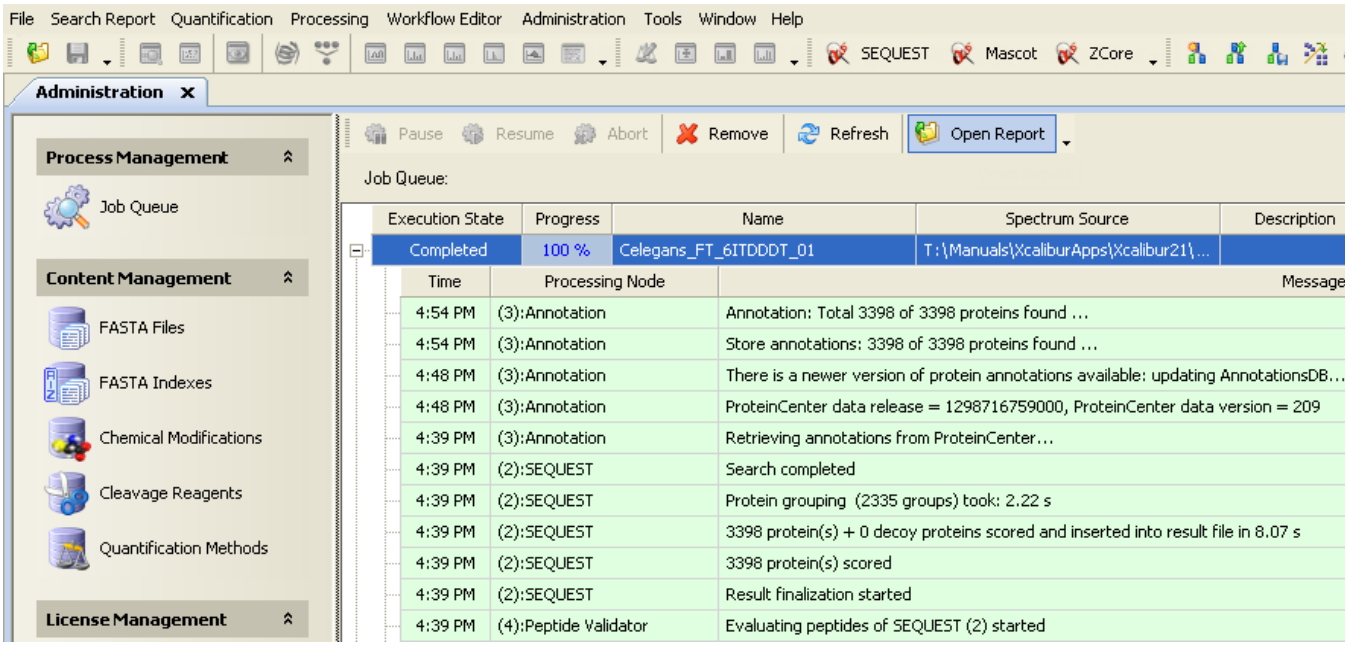
   The example in Figure 13 uses Celegans_FT_6ITDDT_017.

2. **Click** Open Report.

   The Input Files page appears so that you can open the MSF file. For information about opening MSF files, see "Opening MSF Files" on page 215.

> **Note** You can still access the job queue by clicking the Administration tab.

Figure 13. Opening a report in the job queue



Search Results Window Icons

When you apply one of the Show commands on the Search Report menu to the Proteins, Peptides, and Search Input pages, the Proteome Discoverer application opens a search results window. By default, this window opens at the bottom of the page. You can use the icons on this window to manipulate the window.

❖ To hide a results window

- Click and choose **Hide** to remove the search results window.

  You cannot retrieve the window. You must choose the appropriate Show command or icon again to reconstruct the view.

–or–

1. Click and choose **Auto Hide**, or click the **Auto Hide** icon, .

   The search report window disappears but returns when you move the cursor over the tab at the bottom of the window.

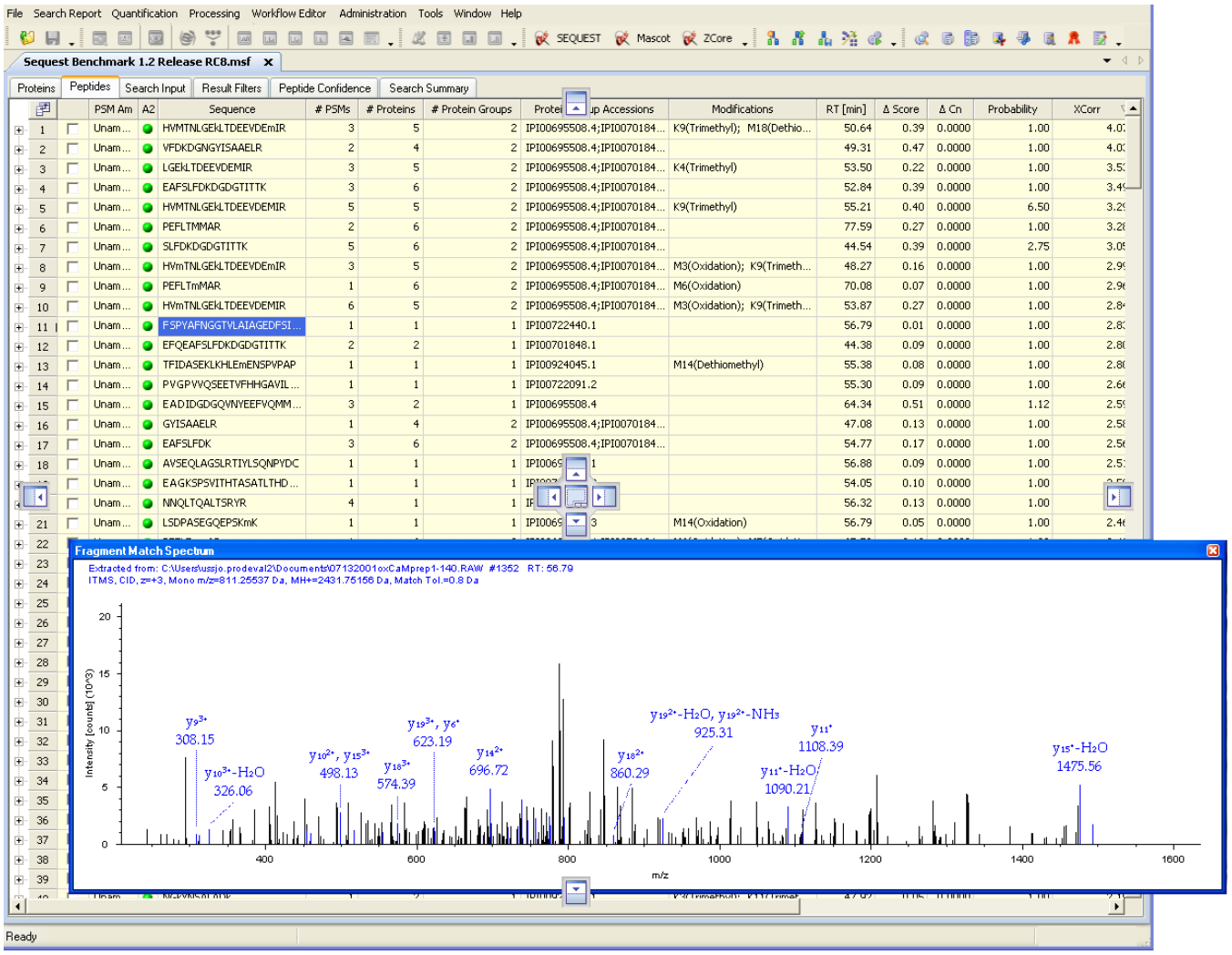
2. Click to turn off the Auto Hide feature.

❖ To move a results window

1. Drag the window by its title bar.

   The group of pointers appears, along with additional pointers at a distance, as shown in Figure 14.

Figure 14. Search results window pointers



2. Drag the window until the cursor aligns with the appropriate pointer. For example, if you want to position the search results window to the right of the results report, drag the search results window until the cursor lines up with the right pointer,  .

❖ To close a results window

- Click  .

Table 4 describes the purpose of the icons that appear in the results windows activated by many Proteome Discoverer application commands.

Table 4. Results window icons

Icon	Function
	Opens a menu with the following three commands: • Hide: Removes the displayed search report window, which you cannot retrieve later. You must choose the appropriate Show command or icon again to reconstruct the view. • Floating: Detaches the search report window from the Proteins, Peptides, or Search Input page. By default, this view appears in the upper left corner of the screen. • Auto Hide: Hides the plot but displays it when you move the cursor over the tab at the bottom of the window.
	Hides the plot but displays it when you move the cursor over the tab.
	Closes the search report window.

Dialog Box Icons

Some Proteome Discoverer application dialog boxes feature the icons shown in Table 5.

Table 5. Dialog box icons

Icon	Description
 Categorized	Categorizes the areas of the dialog box by headings and numbers in the dialog box.
 Alphabetical	Lists the areas of the dialog box alphabetically.

Customizing the User Interface

You can customize the Proteome Discoverer toolbar and the menus.

- Customizing the Toolbar
- Customizing Menus
- Customizing the Job Queue Toolbar
- Setting Global Default Fragment Match Options

Customizing the Toolbar

Most of the application's most commonly used features are accessible through its toolbar icons. Toolbar icons are available for most menu commands. You can temporarily customize the toolbar and also restore it to its default settings.

Note Customized toolbars are restored to their defaults when you restart the application.

- Making Toolbars Visible on the Main Toolbar
- Removing a Single Toolbar Icon from a Toolbar
- Adding a New Toolbar
- Deleting a Toolbar
- Renaming a Toolbar
- Repositioning a Toolbar
- Customizing Toolbar Icons, Fonts, and ToolTips Display

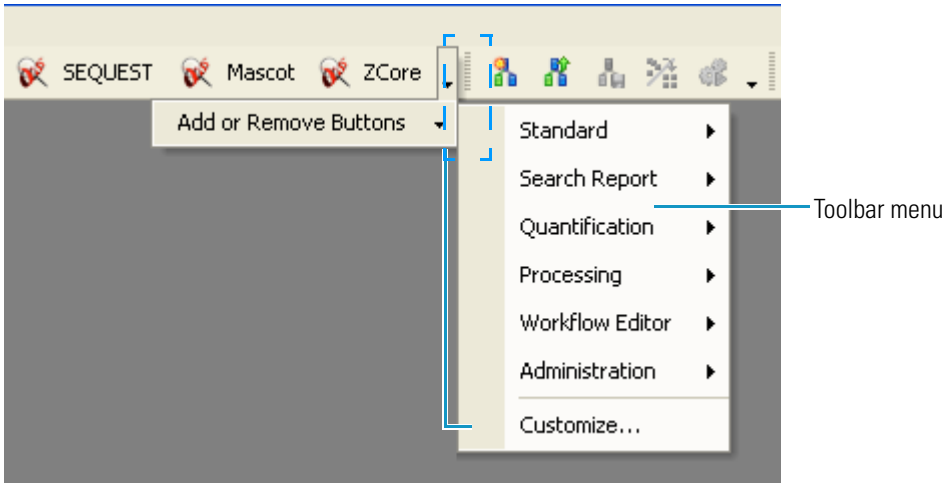
Making Toolbars Visible on the Main Toolbar

You can determine whether to hide or display a toolbar on the main toolbar of the Proteome Discoverer interface.

❖ **To make a toolbar visible in the main toolbar**

1. On the toolbar, click  to access the toolbar menu, shown in Figure 15.

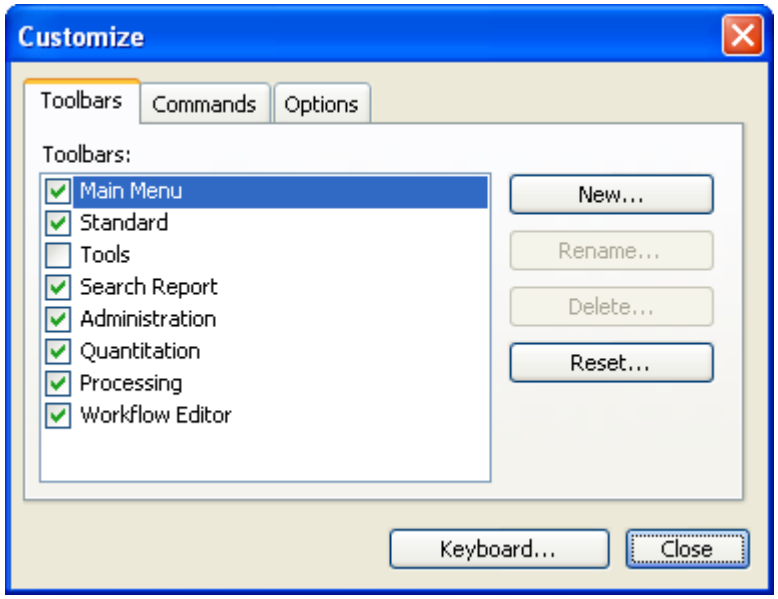
Figure 15. Accessing the toolbar menu



2. Choose **Customize**.

   The Customize dialog box appears, as shown in Figure 16.

Figure 16. Customize dialog box



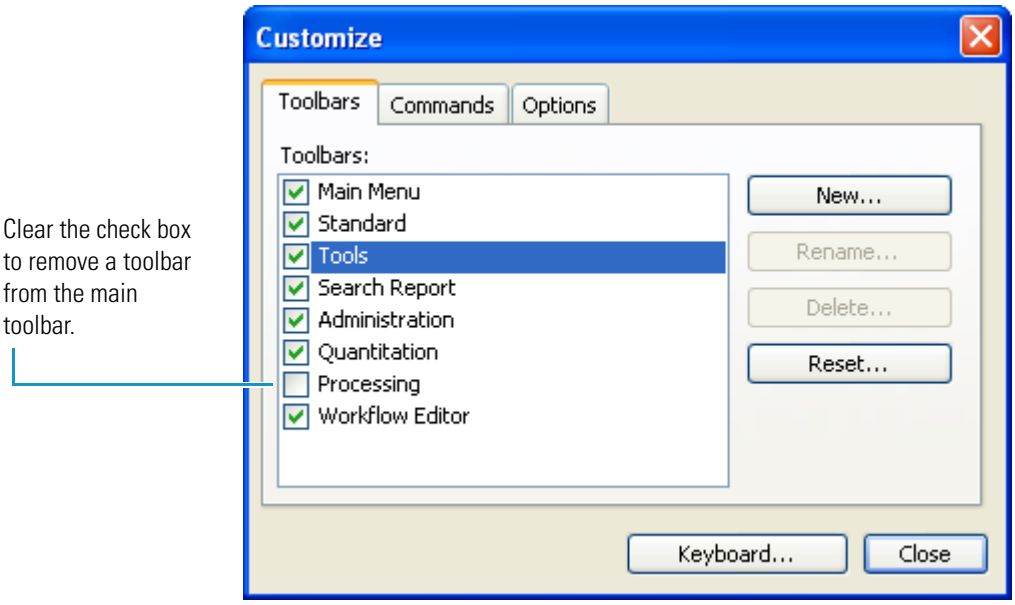
3. Click the **Toolbars** tab if it is not already selected.
4. Select the check box next to the toolbar that you want to make visible on the main toolbar.
5. Click **Close**.

❖ **To remove a toolbar from the main toolbar**

1. On the toolbar, click ▾ to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears.
3. Click the **Toolbars** tab if it is not already selected.
4. Clear the check box next to the toolbar that you want removed from the main toolbar, as shown in Figure 17.

   Figure 17. Removing a toolbar from the main toolbar



   The toolbar disappears from the main toolbar.
5. Click **Close** to close the Customize dialog box.

Removing a Single Toolbar Icon from a Toolbar

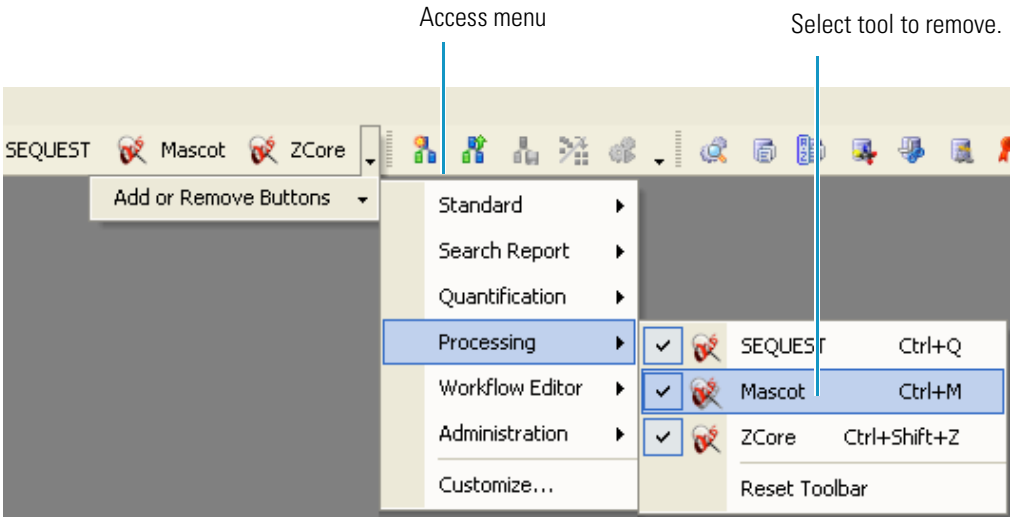
You can remove individual icons from a toolbar.

❖ **To remove a single toolbar icon from a toolbar**

1. On the main toolbar, click ▾ to access the toolbar menu, as shown in Figure 18.

   The active toolbar commands have a check mark next to them.

Figure 18. Removing a tool from the toolbar



2. Clear the check mark adjacent to the tool that you want removed from the toolbar.

   The tool is visible on the menu, but the icon no longer appears on the toolbar.

Adding a New Toolbar

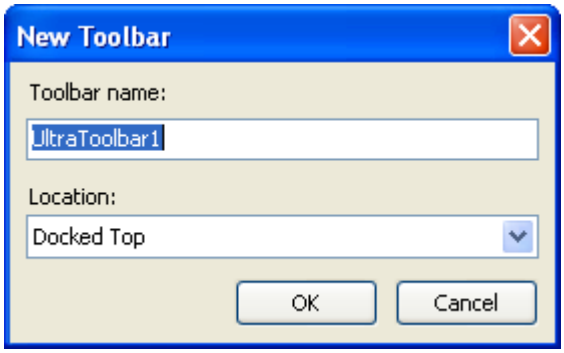
You can create a new toolbar and add it to the Customize dialog box.

❖ To add a new toolbar

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Toolbars** tab, if it is not already selected.
4. Click **New**.

   The New Toolbar dialog box appears, as shown in Figure 19.

Figure 19. New Toolbar dialog box



5. In the Toolbar Name box, type the name for the new toolbar configuration.
6. In the Location list, select the location of the new toolbar:
   - Docked Top: Places the down arrow key, , at the top of the Administration page window. This option is the default.
   - Docked Bottom: Places the down arrow key, , at the bottom of the Administration page window.
   - Docked Left: Places the down arrow key, , at the left of the Administration page window.
   - Docked Right: Places the down arrow key, , at the right of the Administration page window.
   - Floating: Randomly places the down arrow key, , at the top, bottom, left, or right of the Administration page window.
7. Click **OK**.

   The name of the new toolbar is now added to the Toolbars page of the Customize dialog box. To add the new toolbar to the main toolbar, see "Making Toolbars Visible on the Main Toolbar" on page 36.
8. Drag icons from the main toolbar or other toolbar to the location of the new toolbar.

Deleting a Toolbar

You can delete a toolbar that you have added from the Customize dialog box. However, you cannot delete a toolbar that is part of the initial Proteome Discoverer interface.

❖ To delete a toolbar

1. On the toolbar, click to access the toolbar menu.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Toolbars** tab, if it is not already selected.
4. Select the toolbar that you want to delete.
5. Click **Delete**.
6. In the Confirm Toolbar Delete confirmation box, click **OK**.

The name of the customized toolbar is removed from the Toolbars page of the Customize dialog box.

Renaming a Toolbar

You can rename a toolbar that you added by using the Customize dialog box. However, you cannot rename a toolbar that is part of the initial application interface.

❖ **To rename a toolbar**

1. On the toolbar, click to access the toolbar menu.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Toolbars** tab.
4. Select the toolbar that you want to rename.
5. Click **Rename**.
6. In the Rename Toolbar dialog box, type the new name in the Toolbar name box, and click **OK**.

The name of the customized toolbar now appears on the Toolbars page.

Repositioning a Toolbar

❖ **To reposition a toolbar**

1. Select immediately to the left of the toolbar section that you want repositioned.

   A set of cross arrows ( ) appears when you select the toolbar section.
2. Drag the toolbar section to its new position.

Restoring Toolbar Defaults

After you customize the toolbar, you can return it to the original settings.

❖ **To return toolbar settings to their defaults**

1. On the toolbar, click to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears, as shown in Figure 16.
3. Click the **Toolbars** tab.
4. Click **Reset**.
5. In the Confirm Toolbar Reset confirmation box, click **OK**.
6. Click **Close** to close the Customize dialog box.

Customizing Toolbar Icons, Fonts, and ToolTips Display

Use the Options page of the Customize dialog box to customize toolbar features. You can resize the toolbar icons and fonts, and change the ToolTips display to include corresponding icons and shortcut keys or to hide the ToolTips display from view.

- To resize icons on toolbars

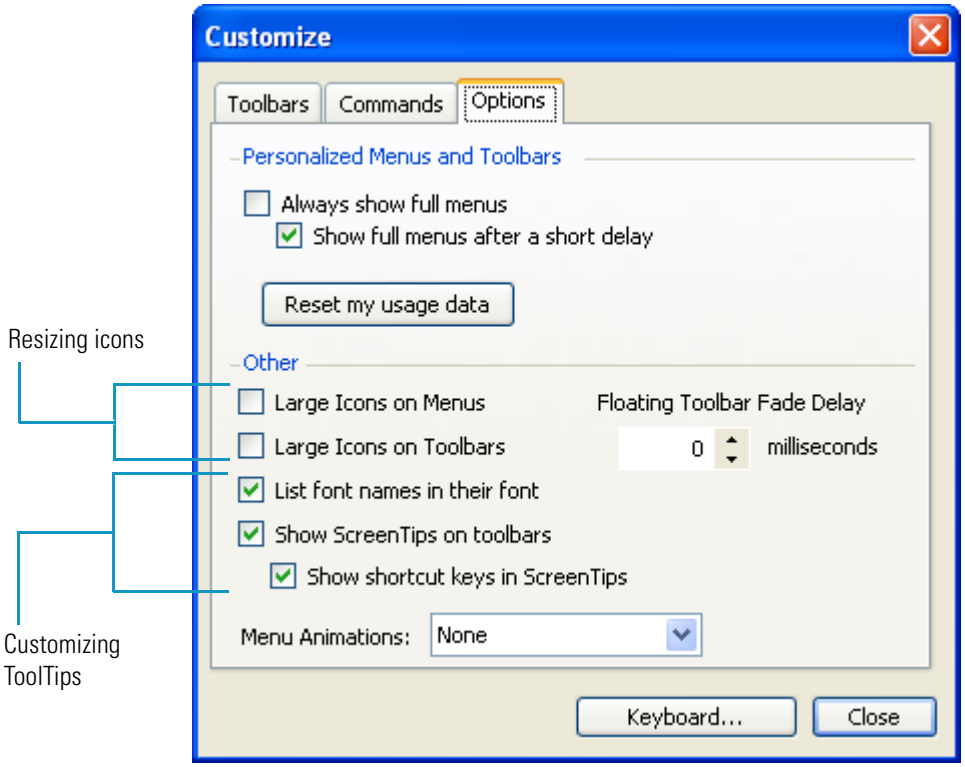
- To modify icons and menu commands
- To customize ToolTips
- To hide ToolTips

❖ To resize icons on toolbars

1. On the toolbar, click ▾ to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears.
3. Click the **Options** tab, shown in Figure 20.

   Figure 20. Customize dialog box



4. To increase the size of the icons on the menus, select the **Large Icons on Menus** option.
5. To increase the size of the icons on the main toolbar, select the **Large Icons on Toolbars** option.
6. To save your changes and close the Customize dialog box, click **Close**.

❖ To modify icons and menu commands

1. On the toolbar, click ▾ to access the toolbar menu.
2. Choose **Customize**.

The Customize dialog box appears.

3. Click the **Commands** tab.
4. Select the icon or menu command.
5. Click **Modify Selection** on the Commands page.
6. Choose an option.

   See the Help in any Microsoft Office product for information about the options available.

7. Click **Close**.

❖ To customize ToolTips

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears.

3. Click the **Options** tab.
4. To display ToolTips when you move the cursor over the icons, select the **Show ScreenTips on Toolbars** option.
5. (Optional) To display shortcut keys with the ToolTips, select the **Show Shortcut Keys in ScreenTips** option.

   You must select Show ScreenTips on Toolbars to use the Show Shortcut Keys in ScreenTips option.

6. To save your changes and close the Customize dialog box, click **Close**.

❖ To hide ToolTips

1. On the toolbar, click the down arrow key,  , to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears.

3. Click the **Options** tab.
4. To hide the ToolTips when you hover the cursor over the icons, clear the **Show ScreenTips on toolbars** option.
5. To save your changes and close the Customize dialog box, click **Close**.

Customizing Menus

You can customize the menus on the Proteome Discoverer application interface by associating them with shortcut keys and specifying how the menus deploy.

This section describes the following functions:

- Adding and Removing Shortcut Keys
- Restoring Shortcut Key Defaults
- Displaying a Command's Function
- Displaying Full or Partial Menus
- Specifying the Menu Deployment Method
- Customizing Dialog Box Parameters
- Customizing Keyboard Dialog Box Parameters

Adding and Removing Shortcut Keys

You can add and remove shortcut keys from commands.

- To add a shortcut key to a command
- To remove a shortcut key from a command

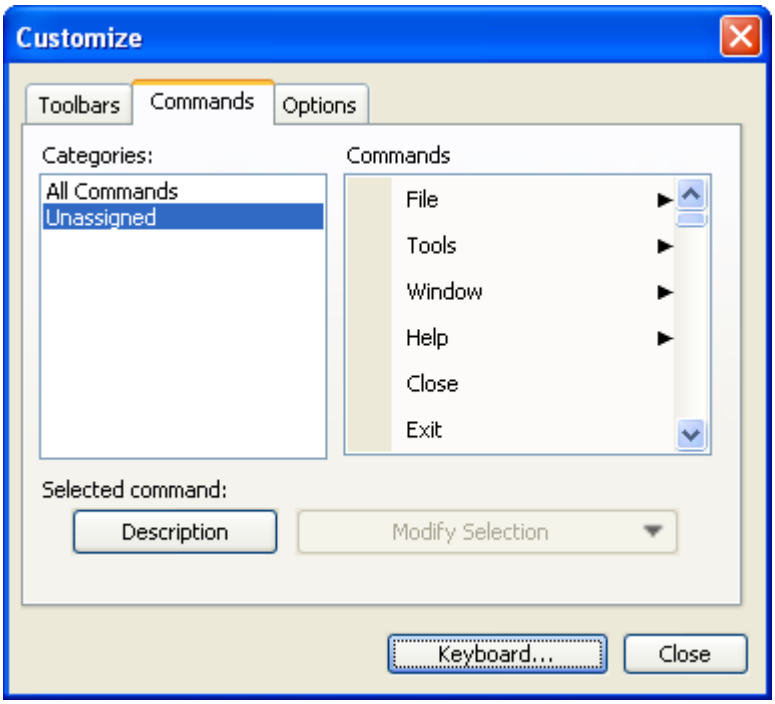
❖ To add a shortcut key to a command

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears.
3. Click the **Commands** page.

   The Commands page appears, as shown in Figure 21.

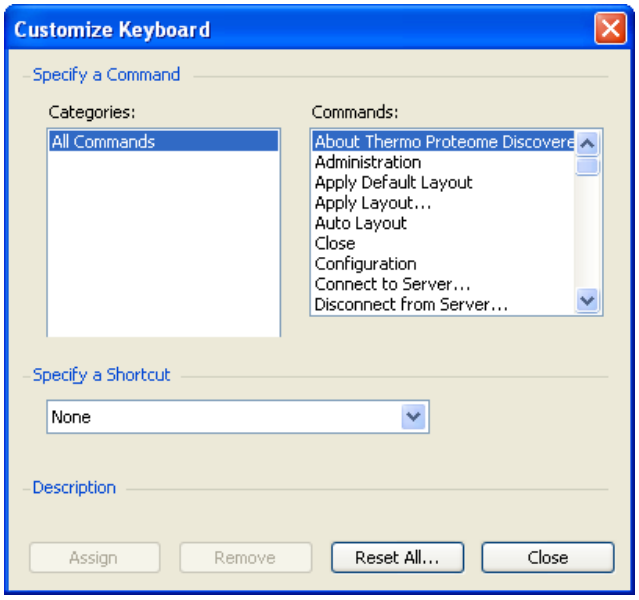
Figure 21. Commands page of the Customize dialog box



4. Click **Keyboard** in the Commands list.

   The Customize Keyboard dialog box opens, as shown in Figure 22.

Figure 22. Customize Keyboard dialog box



5. In the Customize Keyboard dialog box, scroll through the Commands list to find the command and its corresponding keyboard shortcut. From the Specify a Shortcut list, select a shortcut.

6. Below the Specify a Shortcut list, check the message to determine if the shortcut is unassigned.
7. If the shortcut is unassigned, click **Assign**.

   The shortcut is assigned to your command choice.
8. In the Customize Keyboard dialog box, click **Close** to save your changes.
9. In the Customize dialog box, click **Close**.

❖ **To remove a shortcut key from a command**

1. In the Commands list in the Customize Keyboard dialog box, select the command whose shortcut key you want to remove.
2. Click **Remove**.

   The shortcut key is removed from the command.

   If you want to assign a new shortcut key to the command, follow the procedure in "To add a shortcut key to a command" on page 44.
3. In the Customize Keyboard dialog box, click **Close** to save your changes.
4. In the Customize dialog box, click **Close**.

Restoring Shortcut Key Defaults

You can restore the default shortcut keys to commands.

❖ **To restore the default shortcut keys to commands**

1. In the Customize Keyboard dialog box, click **Reset All**.
2. In the Reset Shortcut Customizations confirmation box, click **OK**.
3. In the Customize Keyboard dialog box, click **Close** to save your changes.
4. In the Customize dialog box, click **Close**.

Displaying a Command's Function

You can access context-sensitive Help for information about a command.

❖ **To obtain information about a command's function**

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.

   The Customize dialog box appears.
3. Click the **Commands** page.

4. In the Commands list, click the command that you want information about.
5. Click **Description**.

Displaying Full or Partial Menus

You can determine whether all commands or only the most recently used commands first appear when you select a menu.

❖ To initially display all the commands in a menu

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Options** page.
4. Select the **Always Show Full Menus** check box.
5. Click **Close**.

❖ To initially display the most recently used commands in a menu

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Options** page.
4. Clear the **Always Show Full Menus** check box, if it is selected.
5. Select the **Show Full Menus After a Short Delay** check box.
6. Click **Close**.

❖ To restore the default set of visible commands to the menus and toolbars

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Options** page.
4. Click **Reset My Usage Data**.

Specifying the Menu Deployment Method

You can determine how the menus deploy when you click the menu name.

❖ To specify the menu deployment method

1. On the toolbar, click  to access the toolbar menu.
2. Choose **Customize**.

3. In the Customize dialog box, click the **Options** page.
4. In the Menu Animations box, select an option from the list:
   - Random: Selects the Unfold, Slide, or Fade option at random when you click the menu name.
   - Unfold: Expands menus from the left side of the menu name to the right to fully display them when you click the menu name.
   - Slide: Slides menus down from the menu name to fully display them when you click the menu name.
   - Fade: Gradually infuses menus with color to fully display them when you click the menu name.
   - None: Instantaneously fully displays menus when you click the menu name.

   The default is None.
5. Click **Close**.

Customizing Dialog Box Parameters

The pages of the Customize dialog box contain the parameters described in the following sections.

Toolbars Page

Table 6 describes the parameters on the Toolbars page of the Customize dialog box.

Table 6. Parameters on the Toolbars page of the Customize dialog box (Sheet 1 of 2)

Parameter	Description
Toolbars	Lists the toolbars on the Proteome Discoverer application interface and determines whether they are visible on the main toolbar of the interface.
New	Opens the New Toolbar dialog box so that you can add a new toolbar.
Rename	Opens the Rename Toolbar dialog box so that you can assign the toolbar a new name.
Delete	Deletes the selected toolbar from the Toolbars page of the Customize dialog box.
Reset	Returns the toolbar settings to their defaults.

Table 6. Parameters on the Toolbars page of the Customize dialog box (Sheet 2 of 2)

Parameter	Description
Keyboard	Opens the Customize Keyboard dialog box so that you can assign a shortcut key to a command. For a description of the parameters on the Customize Keyboard dialog box, see "Customizing Keyboard Dialog Box Parameters" on page 50.
Close	Closes the Toolbars page.

Commands Page

Table 7 describes the parameters on the Commands page of the Customize dialog box.

Table 7. Parameters on the Commands page of the Customize dialog box

Parameter	Description
Categories	Lists the type of commands that appear in the Commands box of the Commands page.
All Commands	Lists all commands.
Unassigned	Lists all the commands that are not assigned to a toolbar.
Commands	Lists the commands selected in the Categories box of the Commands page.
Description	Opens the Help topic for the selected command.
Keyboard	Opens the Customize Keyboard dialog box so that you can assign a shortcut key to a command.
Close	Closes the Commands page.

Options Page

Table 8 describes the parameters on the Options page of the Customize dialog box.

Table 8. Parameters on the Options page of the Customize dialog box (Sheet 1 of 2)

Parameter	Description
Always show full menus	Initially displays all the commands in a menu when you select the menu.
Show full menus after a short delay	Initially displays only the most recently used commands in a menu when you select the menu.
Reset my usage data	Restores the default set of visible commands to the menus and toolbars.
Large Icons on Menus	Increases the size of the icons on the menus.
Large Icons on Toolbars	Increases the size of the icons on the toolbars.

Table 8. Parameters on the Options page of the Customize dialog box (Sheet 2 of 2)

Parameter	Description
Show ScreenTips on toolbars	Determines whether ToolTips explaining the function of a toolbar icon appear when you move the cursor over the icon.
Show shortcut keys in ScreenTips	Determines whether shortcut keys are displayed in the ToolTips that appear when you select the Show ScreenTips option on toolbars.
Menu Animations	Determines how the menus deploy when you click the menu names. • Random: Selects the Unfold, Slide, or Fade option at random when you click the menu name. • Unfold: Expands menus from the left side of the menu name to the right to fully display them when you click the menu name. • Slide: Slides menus down from the menu name to fully display them when you click the menu name. • Fade: Gradually infuses menus with color to fully display them when you click the menu name. • None (Default): Instantaneously fully displays menus when you click the menu name.
Keyboard	Opens the Customize Keyboard dialog box, shown in Figure 22 on page 45, so that you can assign keyboard shortcuts to commands.
Close	Closes the Customize dialog box

Customizing Keyboard Dialog Box Parameters

Table 9 describes the parameters on the Customize Keyboard dialog box.

Table 9. Parameters on the Customize Keyboard dialog box (Sheet 1 of 2)

Parameter	Description
Categories	Lists the categories of commands.
Commands	Lists all the commands on the interface menus.
Specify a Shortcut	Specifies the shortcut to assign to the selected command.
Assign	Assigns the shortcut selected in the Specify a Shortcut list to the selected command in the Commands box.

Table 9. Parameters on the Customize Keyboard dialog box (Sheet 2 of 2)

Parameter	Description
Remove	Removes the shortcut key from the selected command in the Commands box.
Reset All	Restores the shortcut key defaults.
Close	Closes the Customize Keyboard dialog box.

Customizing the Job Queue Toolbar

You can choose the icons that appear on the toolbar of the job queue. The default icons are the Pause, Resume, Abort, Remove, Refresh, and Open Report icons.

- To display or hide the job queue toolbar
- To customize the job queue toolbar
- To add a new job queue toolbar
- To rename a job queue toolbar
- To delete a job queue toolbar

❖ To display or hide the job queue toolbar

1. Click  to the right of the Open Report icon.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Toolbars** tab, if it is not already selected.
4. Clear the **UltraToolbarJobQueue** check box to hide the job queue toolbar.
5. Click **Close**.
6. To redisplay the toolbar, close the Administration tab and reopen it.

❖ To customize the job queue toolbar

1. Click  to the right of the Open Report icon.
2. Choose **UltraToolbarJobQueue**, and on the menu that appears, clear or select the check boxes corresponding to the appropriate icons.
3. To restore all the icons to the toolbar, choose **Reset Toolbar** from the menu.

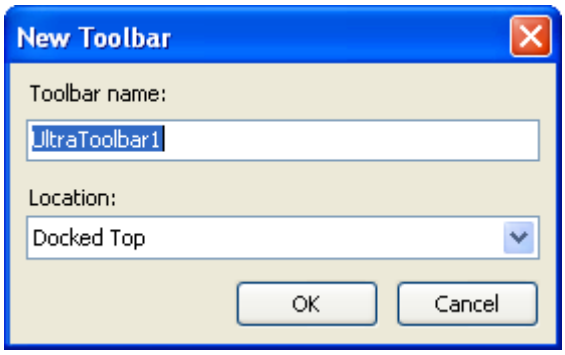
❖ To add a new job queue toolbar

1. Click  to the right of the Open Report icon.
2. Choose **Customize**.

3. In the Customize dialog box, click the **Toolbars** tab, if it is not already selected.
4. Click **New**.

   The New Toolbar dialog box appears, as shown in Figure 23.

   Figure 23. New Toolbar dialog box



5. In the Toolbar Name box, type the name for the new toolbar configuration.
6. In the Location list, select the location of the new toolbar menu:
   - Docked Top: Places the down arrow key, , at the top of the Administration page window. This option is the default.
   - Docked Bottom: Places the down arrow key, , at the bottom of the Administration page window.
   - Docked Left: Places the down arrow key, , at the left of the Administration page window.
   - Docked Right: Places the down arrow key, , at the right of the Administration page window.
   - Floating: Randomly places the down arrow key, , at the top, bottom, left, or right of the Administration page window.
7. Click **OK**.

   The name of the new toolbar is now added to the Toolbars page of the Customize dialog box.
8. Drag icons from the main toolbar or other toolbar to the location of the new toolbar.

❖ To rename a job queue toolbar

1. Click the down arrow key, , to the right of the Open Report icon.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Toolbars** tab, if it is not already selected.
4. In the list of toolbars on the Toolbars page, select the name of the toolbar that you want to rename.

5. Click **Rename**.
6. In the Rename Toolbar dialog box, type over the existing toolbar name with the new name.
7. Click **OK**.

❖ To delete a job queue toolbar

1. Click  to the right of the Open Report icon.
2. Choose **Customize**.
3. In the Customize dialog box, click the **Toolbars** tab, if it is not already selected.
4. In the list of toolbars on the Toolbars page, select the name of the toolbar that you want to delete.
5. Click **Delete**.
6. In the Confirm Toolbar Delete confirmation box, click **OK**.

The name of the toolbar is now deleted from the list of toolbars on the Toolbars page.

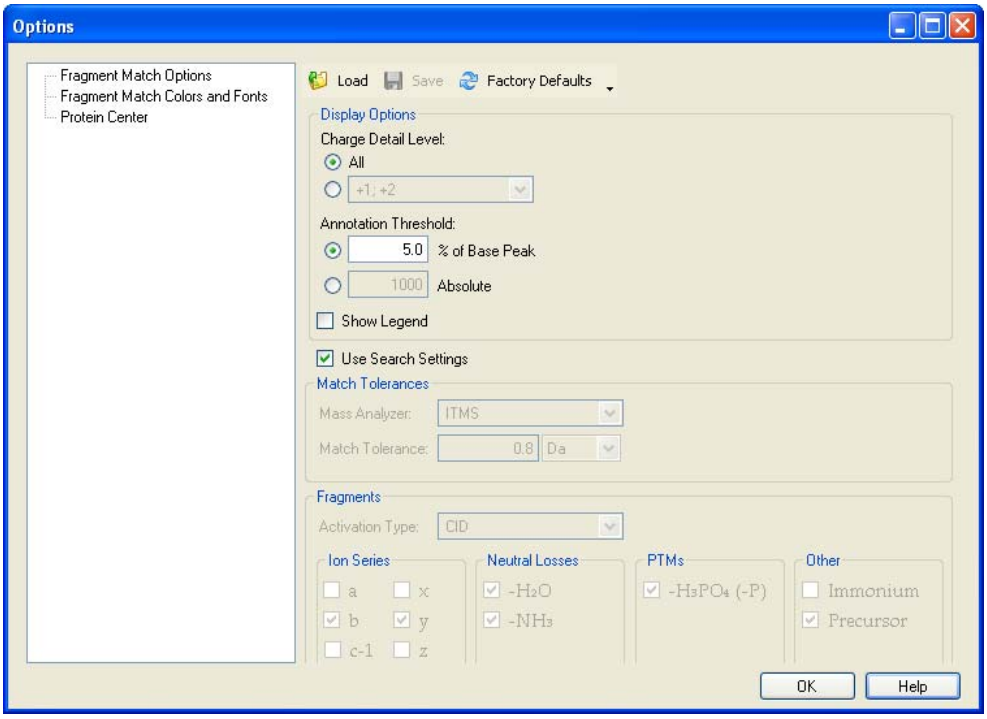
Setting Global Default Fragment Match Options

You can globally modify the default colors and fonts that the Proteome Discoverer application uses to annotate fragment peaks in spectra. The ion tables in the Peptide Identification Details dialog box also use these colors. When you save these settings, they become the global default settings for a particular mass analyzer/activation type combination. If necessary, you can restore the factory default settings, which use the most common ion series, match tolerance, neutral losses, and PTMs for the selected mass analyzer/activation type combination.

❖ To set the global default fragment match options

1. Choose **Tools > Options** to display the Fragment Match Options page of the Options dialog box, shown in Figure 24.

Figure 24. Fragment Match Options page of the Options dialog box

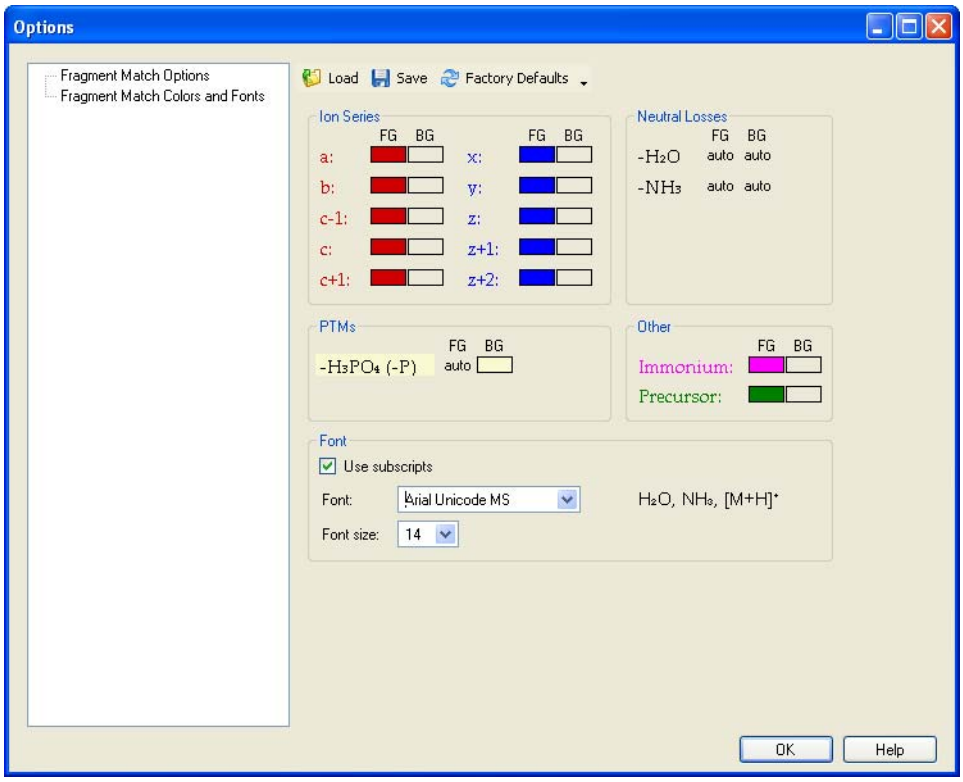


2. Click **Fragment Match Colors and Fonts** in the Options tree.

   The Fragment Match Colors and Fonts page of the Options dialog box is shown in Figure 25.

Note The Proteome Discoverer application automatically chooses the fragment match colors of PTMs and neutral loss ions on the basis of the colors of the ion series that they belong to.

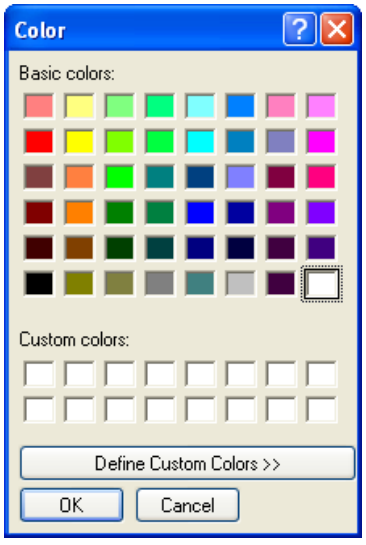
Figure 25. Fragment Match Colors and Fonts page of the Options dialog box



3. Click the foreground (FG) or background (BG) color bar next to the ion whose match color you want to change.

   The Color dialog box appears, as shown in Figure 26.

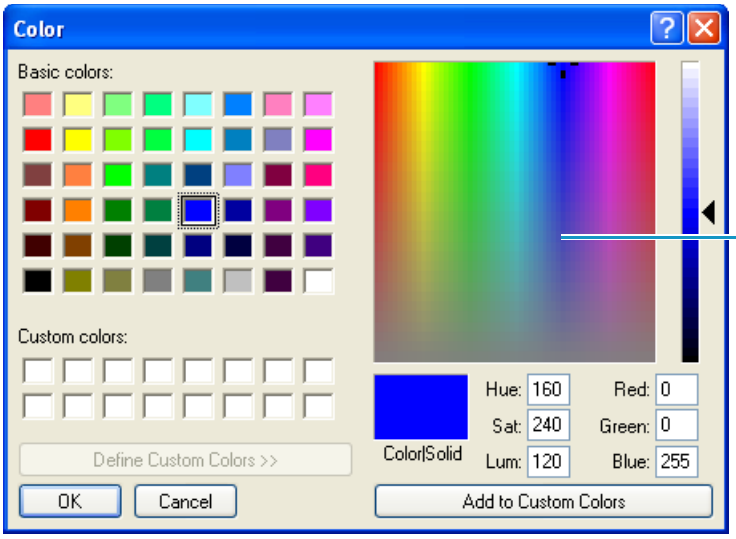
Figure 26. Color dialog box



4. To create a custom color:
   a. Click **Define Custom Colors** to display the custom colors palette.

      The custom colors palette appears, as shown in Figure 27.

      Figure 27. Custom colors palette in the Color dialog box



   b. Select a color in the custom colors palette.
   c. Click **Add to Custom Colors** to add the color to the Custom Colors area.
5. Select a color:
   - For a basic color, select a color in the Basic Colors area.
   - For a custom color, select a color in the Custom Colors area.
6. Click **OK**.

Options Dialog Box Parameters

Table 10 lists the parameters in the Options dialog box.

Table 10. Options dialog box parameters (Sheet 1 of 2)

Command or Option	Description
Fragment Match Options	Opens the Fragment Match Options page shown in Figure 24 on page 54. For information on the parameters on this page, see Table 11.

Table 10. Options dialog box parameters (Sheet 2 of 2)

Command or Option	Description
Fragment Match Colors and Fonts	Opens the Fragments Match Colors and Fonts page shown in Figure 25 on page 55. For information on the parameters on this page, see Table 12.
ProteinCenter	Opens the ProteinCenter page shown in Figure 237 on page 446. For information on the parameters on this page, see Table 69.

The following sections describe the parameters on these pages of the Options dialog box:

- Fragment Match Options Page Parameters
- Fragment Match Colors and Fonts Page Parameters
- ProteinCenter Page Parameters

Fragment Match Options Page Parameters

Table 11 describes the parameters on the Fragment Match Options page of the Options dialog box.

Table 11. Fragment Match Options page parameters (Sheet 1 of 2)

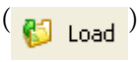
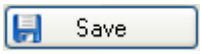
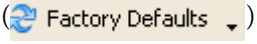
Command or Option	Description
Load ()	Loads the last saved default fragment match settings for the selected mass analyzer/activation type combination.
Save ()	Saves the current fragment match settings as the default fragment match settings for the selected mass analyzer/activation type combination.
Factory Defaults ()	Loads the original factory fragment match settings for all mass analyzer/activation type combinations.
Charge Detail Level	Selects the target charge state to display. You can choose from the following settings: • (Default) All: Displays all fragment ions, regardless of charge. • +1: Displays singly charged fragment ions. • +2: Displays doubly charged fragment ions. • +3: Displays triply charged fragment ions. • +4 …: Displays additional charged fragment ions if available.

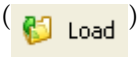
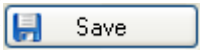
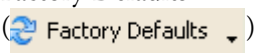
Table 11. Fragment Match Options page parameters (Sheet 2 of 2)

Command or Option	Description
Annotation Threshold	Specifies a peak intensity limit. Peaks below this limit are not annotated. This limit prevents the addition of too many labels to a spectrum, especially if the low-intensity peaks are possibly noise peaks. • (Default) % of Base Peak: Specifies an intensity threshold as a percentage of the base peak in the spectrum. The base peak is the largest peak in a spectrum. • Absolute: Specifies an absolute intensity threshold.
Show Legend	Determines whether a legend consisting of colored lines that denote the displayed ion series appears at the bottom of each spectrum image.
Use Search Settings	Annotates the spectra using the Match Tolerances and Fragments settings used for the search.
Mass Analyzer	Specifies which mass analyzer was used to obtain the spectrum.
Mass Tolerance	Specifies the mass tolerance. The Proteome Discoverer application annotates the spectrum with ions that it assigns within this mass tolerance. You can use daltons or millimass units.
Activation Type	Specifies what type of activation produced the fragments.
Ion Series	Selects which types of ion series to annotate in the spectra.
Neutral Losses	Selects which types of neutral losses to annotate in the spectra.
PTMs	Annotates the spectra ions that have lost H_3PO_4. This loss is represented as –P in the ion list.
Other	Annotates the spectra with immonium ions or precursor ions.

Fragment Match Colors and Fonts Page Parameters

Table 12 lists the parameters in the Fragment Match Colors and Fonts page of the Options dialog box.

Table 12. Fragment Match Colors and Fonts page parameters

Command or Option	Description
Load ()	Loads the last saved default fragment match settings for the selected mass analyzer/activation type combination.
Save ()	Saves the current fragment match settings as the default fragment match settings for the selected mass analyzer/activation type combination.
Factory Defaults ()	Loads the original factory fragment match settings for all mass analyzer/activation type combinations.
Ion Series	Specifies the foreground and background colors of the fragment ions found in the search.
Neutral Losses	The Proteome Discoverer application automatically chooses the fragment match colors of neutral loss ions on the basis of the colors of the ion series that they belong to. The options in the Neutral Losses area are read-only.
PTMs	The Proteome Discoverer application automatically chooses the fragment match colors of PTMs on the basis of the colors of the ion series that they belong to. However, you can select the background color by clicking the BG box.
Other	Specifies the foreground and background colors of any adduct ions found in the search.
Font	Specifies the font used to display annotations and views: • Use Subscripts: Specifies whether subscripts are used in ion names. By default, the Proteome Discoverer application uses subscripts. • Font: Specifies the type of font used to display annotations and views. The default type is Palatino Linotype. • Font Size: Specifies the size of the font used to display annotations and views. The default size is 14.

ProteinCenter Page Parameters

For information on the parameters of the ProteinCenter page of the Options dialog box, see "ProteinCenter Page Parameters" on page 448.

Getting Started

This chapter describes how to use Proteome Discoverer search wizards and its Workflow Editor to define your search parameters. The search wizards are the quickest way to start using the Proteome Discoverer application.

Contents

- Starting the Proteome Discoverer Application
- Closing the Proteome Discoverer Application
- Starting a New Search by Using the Search Wizards
- Starting a New Search by Using the Workflow Editor

Starting the Proteome Discoverer Application

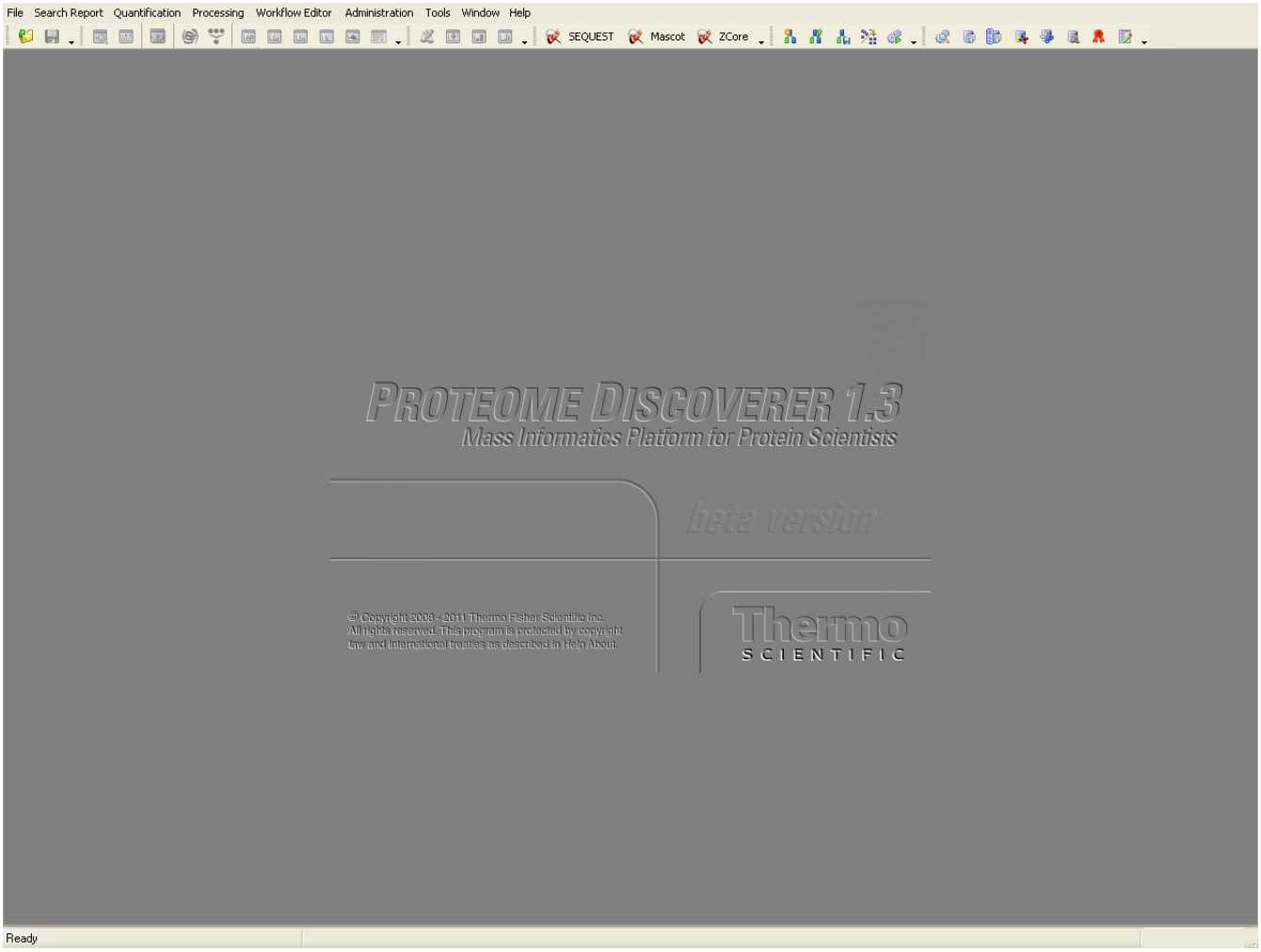
Open the Proteome Discoverer application by choosing a Start menu command or clicking a desktop icon.

❖ **To start the Proteome Discoverer application**

- From the Start menu, choose **Programs > Thermo Proteome Discoverer** or click the **Proteome Discoverer icon**,  , on your desktop.

  The Proteome Discoverer main window opens, as shown in Figure 28.

Figure 28. Proteome Discoverer main window



For information on the features of this window, see "Proteome Discoverer User Interface" on page 19. For instructions on opening an MSF file, see "Opening MSF Files" on page 215.

Closing the Proteome Discoverer Application

Save your changes before you exit the Proteome Discoverer application, because it does not prompt you.

❖ **To close the Proteome Discoverer application**

- Choose **File > Exit**.

  The Proteome Discoverer application closes.

Starting a New Search by Using the Search Wizards

As mentioned, the quickest way to begin using the Proteome Discoverer application is to define your search parameters using the search wizards. You can access three search wizards from the Proteome Discoverer application interface: SEQUEST, Mascot, and ZCore. Use these search wizards to set the search parameters, select a database and a search engine, and select the chemical modifications that you will use to conduct your search.

Note You can use the ZCore search wizard only with mass spectrometers that use the ETD method of peptide fragmentation, such as Orbitrap™ with ETD or LTQ™ XL™ with ETD.

For detailed information about these three wizards, see "Search Engines" on page 3.

You can also access SEQUEST, Mascot, and ZCore through the Workflow Editor.

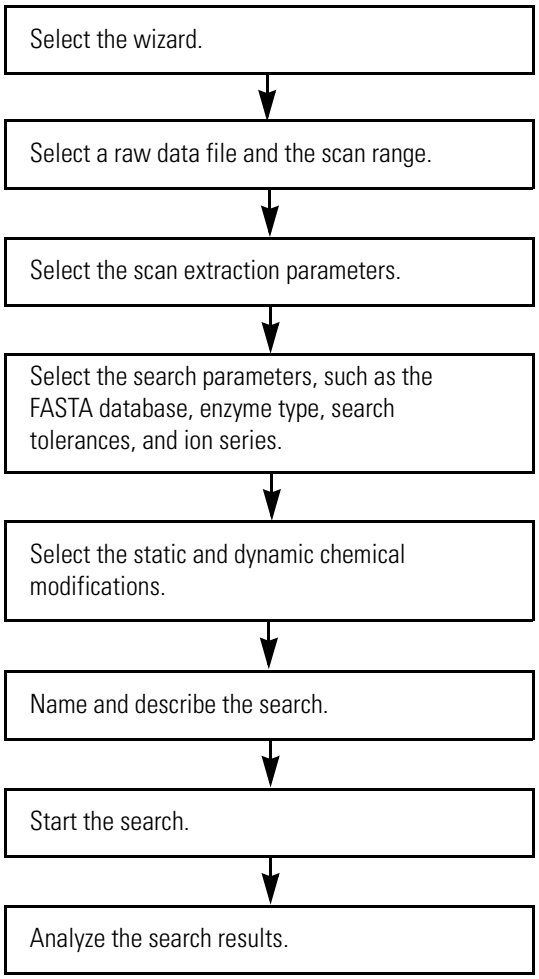
When using the search wizards, you must define two key settings:

- Spectrum source file: raw, MGF, MZDATA, MZXML, or MZML. The search wizards do not support multiple-spectrum source files. To load multiple-spectrum source files, you must use the Workflow Editor. For detailed information about this process, see "Starting a New Search by Using the Workflow Editor" on page 89.
- FASTA file

You can also set dynamic and static chemical modifications.

Figure 29 shows the procedure for using the search engine wizards.

Figure 29. The Proteome Discoverer search wizard process



Configuring Search Parameters

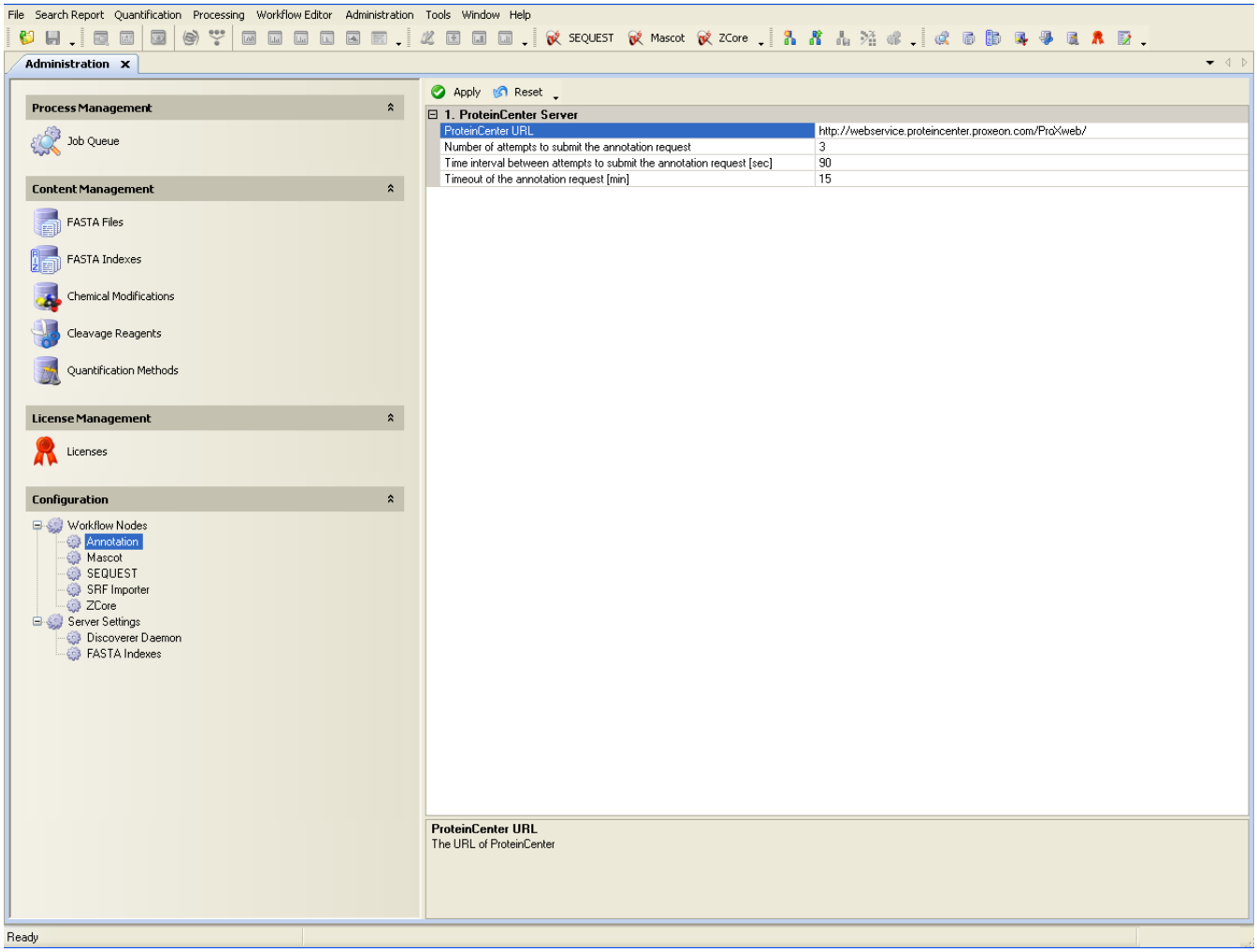
Before you begin the search, you can configure certain search parameters for the SEQUEST, Mascot, and ZCore search engines.

❖ To configure search parameters

1. Choose **Administration > Configuration**, or click the **Edit Configuration** icon, .

   The Administration page changes to the Configuration view, shown in Figure 30.

Figure 30. Configuration view of the Administration page



Configuring the SEQUEST Search Engine

For searches with the SEQUEST search engine, specify how to display the peptide confidence by default. SEQUEST scores the number of fragment ions that are common to two different peptides with the same precursor mass and calculates the cross-correlation score for all candidate peptides queried from the database. By default, it sorts the resulting XCorr value in descending order.

❖ To configure SEQUEST

1. On the Administration page, click **SEQUEST** under Workflow Nodes in the Configuration section.
2. If you are using SEQUEST to search low-resolution data, set the XCorr confidence thresholds under the XCorr Confidence Thresholds (low-resolution data) parameter.

   The default values appear in Figure 31.

Figure 31. XCorr confidence thresholds for SEQUEST

1. XCorr Confidence Thresholds (low-resolution data)	
z=1: High Confidence XCorr	1.5
z=1: Medium Confidence XCorr	0.7
z=2: High Confidence XCorr	2
z=2: Medium Confidence XCorr	0.9
z=3: High Confidence XCorr	2.5
z=3: Medium Confidence XCorr	1.2
z>=4: High Confidence XCorr	3
z>=4: Medium Confidence XCorr	1.5
2. XCorr Confidence Thresholds (high-resolution data)	
z=1: High Confidence XCorr	1.2
z=1: Medium Confidence XCorr	0.7
z=2: High Confidence XCorr	1.9
z=2: Medium Confidence XCorr	0.8
z=3: High Confidence XCorr	2.3
z=3: Medium Confidence XCorr	1
z>=4: High Confidence XCorr	2.6
z>=4: Medium Confidence XCorr	1.2

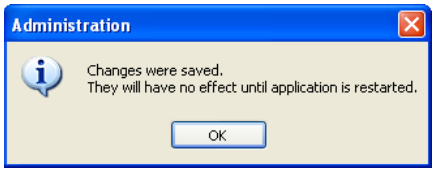
3. If you are using SEQUEST to search high-resolution data, set the XCorr confidence thresholds under the XCorr Confidence Thresholds (high-resolution data) parameter.

   The default values appear in Figure 31.

4. If you changed any settings, click Apply.

   The message box shown in Figure 32 appears:

   Figure 32. Administration message box

5. Click **OK**.

   Note Click Reset to return to the default values.

6. Restart your machine.

Configuring the Mascot Search Engine

Before using the Mascot search engine, you must direct the Proteome Discoverer application to the location of the Mascot server and configure the parameters that control access to the Mascot server. If your Mascot search fails, the following guidelines can help you check for server problems.

Directing the Proteome Discoverer Application to the Mascot Server Location

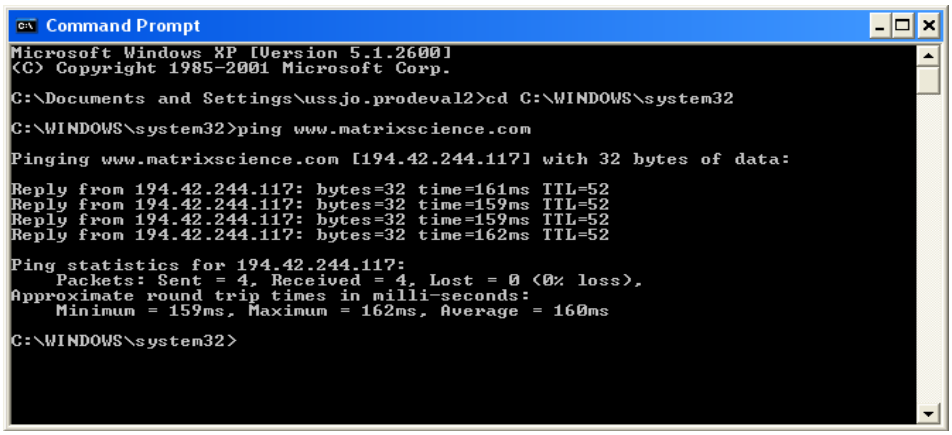
To connect to a Mascot server, refer to the "How to Connect to a Mascot Server" section of the Proteome Discoverer release notes included on every Proteome Discoverer installation DVD. To test the connection between the Proteome Discoverer application and the Mascot server, refer to "Testing the Connection to the Mascot Server," in the *Proteome Discoverer Installation Guide*.

Open a Web browser and try to access the Mascot server through its URL. If you cannot, the Mascot server might not be running, or the URL might not be correct. In this case, contact your system administrator to assist you.

If you can obtain Web access to the Mascot server, you might not be able to reach the server by using the ping command. A firewall on your computer, the Mascot server computer, or a bad network connection might be blocking the ping command.

To find out if the ping command is blocked, open a command shell and type `ping` *`Mascot_server_name`*. If the ping command is successful, the output should resemble that shown in Figure 33. If the ping command is unsuccessful, contact your system administrator to assist you in resolving this problem.

Figure 33. Output of a successful ping command



If you can obtain Web access to the Mascot server and the ping test is successful but the same URL is not accepted in the Proteome Discoverer application, a type of user authentication restriction might be active. In this case, the error message issued by the Proteome Discoverer application should provide information about missing authentication. If it does not, send an error report.

Configuring Mascot Parameters

Before using the Mascot search engine, set the parameters that govern access to the Mascot server.

❖ To configure Mascot

1. On the Administration page, click **Mascot** under Workflow Nodes in the Configuration section.

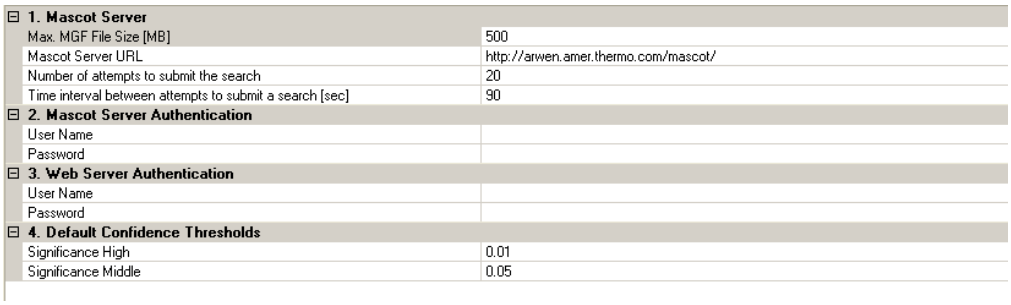
   The Proteome Discoverer application generates an MGF file that contains the search settings and all mass spectral information. It submits this file to the Mascot server through a Web server, which might have a file size limitation. A search that generates large amounts of data—for example, a search with multiple raw files—could create an MGF file that exceeds this limitation. The Max. MGF File Size parameter avoids this limitation by performing several separate Mascot searches and merging the results to avoid this file-size limitation.

2. To split the MGF file and avoid any potential file-size limitations on the Web server, enter the maximum size that the MGF file can be in the Max. MGF File Size [MB] box, in megabytes, as shown in Figure 34.

   This size should be less than the file size permitted by the Web server.

   The minimum file size is 20, and there is no maximum. The default file size is 500 megabytes.

   Figure 34. Maximum MGF file size for Mascot

1. Mascot Server	
Max. MGF File Size [MB]	500
Mascot Server URL	http://arwen.amer.thermo.com/mascot/
Number of attempts to submit the search	20
Time interval between attempts to submit a search [sec]	90
2. Mascot Server Authentication	
User Name	
Password	
3. Web Server Authentication	
User Name	
Password	
4. Default Confidence Thresholds	
Significance High	0.01
Significance Middle	0.05

3. In the Number of Attempts to Submit the Search box, specify the number of times that the Proteome Discoverer application tries to submit the search when the Mascot server is busy.

   The minimum value is 0, and there is no maximum value. The default is 20.

4. In the Time Interval between Attempts to Submit a Search [sec] box, specify the interval of time, in seconds, that elapses between attempts to submit a search when the Mascot server is busy.

   The minimum value is 20, and there is no maximum value. The default is 90 seconds.

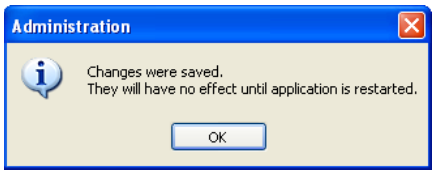
5. If you are accessing a Mascot server through your own network and security for that server is enforced, enter your user name and password in the boxes beneath the Mascot Server Authentication parameter.

6. If you are accessing a Mascot server through the Web and security for that server is enforced, enter your user name and password in the boxes beneath the Web Server Authentication parameter.

7. Set the Default Confidence Thresholds parameters:
   - Significance High: Calculates the thresholds for high -confidence peptides. The Proteome Discoverer application automatically sets this value to the calculated relaxed significance when it performs a decoy search. The minimum value is 0.0, and the maximum value is 1.0. The default is 0.01.
   - Significance Middle: Calculates the thresholds for medium-confidence peptides. The Proteome Discoverer application automatically sets this value to the calculated relaxed significance when it performs a decoy search. The minimum value is 0.0, and the maximum value is 1.0. The default is 0.05.
8. If you changed any settings, click Apply.

   The message box shown in Figure 32 appears:

   Figure 35. Administration message box

9. Click **OK**.

   Note Click Reset to return to the default values.

10. Restart your machine.

Troubleshooting Failed Mascot Searches

If all your searches with Mascot fail, follow these guidelines:

- Verify that the Mascot server is running and accessible from the computer that is running the Proteome Discoverer application. For details on how to do this, see "Directing the Proteome Discoverer Application to the Mascot Server Location" on page 67.
- If your Mascot server is running, verify that it is operating properly by submitting a simple search from the Mascot Web interface. If it is not operating properly, there might be a problem with the Mascot server itself. Contact your system administrator to assist you.
- If your Mascot server is operating properly and you can access it from the Proteome Discoverer application, try to perform a very simple search using the Mascot wizard.
  - If simple searching fails, there might be a general problem in the interaction between the Proteome Discoverer application and the Mascot server. In this case, file an error report.

- If you can perform simple Mascot searches, investigate your failing searches more closely:
  - Does the search finish successfully on the Mascot server according to the Mascot search log?
  - Do the process messages sent to the job queue during the search indicate the problem?

If these suggestions still do not resolve the search problem, file an error report.

Configuring the ZCore Search Engine

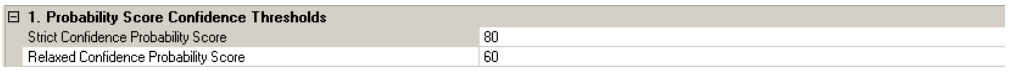
For searches with the ZCore search engine, specify how to display the peptide confidence by default.

❖ To configure ZCore

1. On the Administration page, click **ZCore** under Workflow Nodes in the Configuration section.
2. For the Probability Score Confidence Thresholds parameter, set the Strict Confidence Probability Score and the Relaxed Confidence Probability Score, as shown in Figure 36.

   The default of the Strict Confidence Probability Score is 80, and the default of the Relaxed Confidence Probability Score is 60.

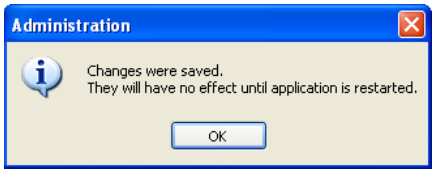
   Figure 36. Probability score confidence thresholds for ZCore

3. If you changed any settings, click Apply.

   The message box shown in Figure 32 appears:

   Figure 37. Administration message box

4. Click **OK**.

   Note Click Reset to return to the default values.

5. Restart your machine.

Starting a New Search

The following procedure describes how to search your data by using a search wizard, using SEQUEST as the example. The procedure is very similar for Mascot and ZCore searches; differences from the SEQUEST procedure are noted where appropriate.

Note Although the basic procedure for using the Mascot wizard and the SEQUEST wizard is the same, see "Configuring the Mascot Search Engine" on page 66 for information about the unique aspects of conducting Mascot searches.

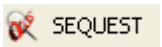
If you have not selected a FASTA database to search, you must add one before you start a search wizard. For instructions on adding a FASTA file, see "Adding FASTA Files" on page 162.

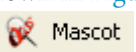
Note The available FASTA files are registered and available through the Proteome Discoverer application. See "Using FASTA Databases" on page 159.

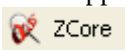
❖ To start a new search using a search wizard

1. (Optional) Open the job queue by choosing **Administration > Show Job Queue** or clicking the **Show Job Queue** icon,  .

   You can find more information about the job queue in "Job Queue" on page 28.

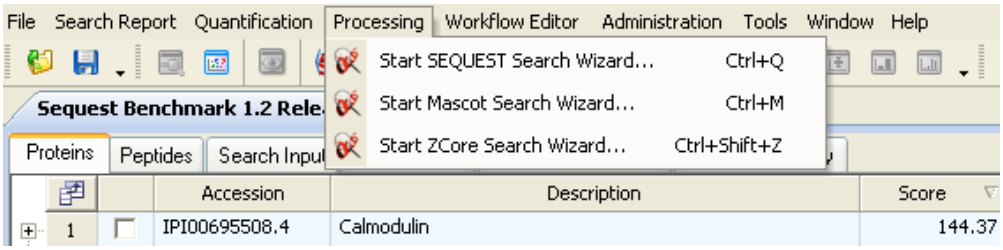
2. Choose **Processing > Start *Wizard_name* Search Wizard**, as shown in Figure 38, or click the appropriate wizard icon: 

for SEQUEST, 

for Mascot, or 

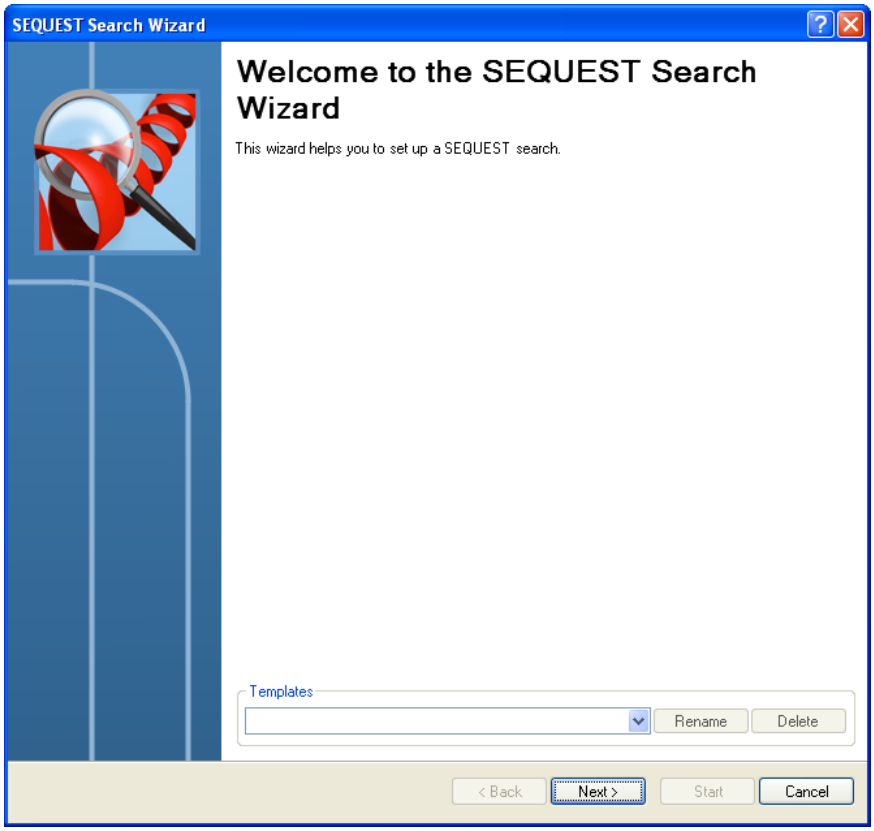
for ZCore.

   Figure 38. Three wizard options in the Processing menu

   The Welcome to the *Wizard_name* Search Wizard page appears, as shown in Figure 39.

Figure 39. Welcome to the *Wizard_name* Search Wizard page



3. To use a template from a previous search, select it from the Templates list.

   To give the selected template a new name, click **Rename**, and in the Renaming Template dialog box, type the new name in the New Name box and click **OK**.

   To delete the selected template, click **Delete** and in the confirmation box, click **OK**.

4. Click **Next**.

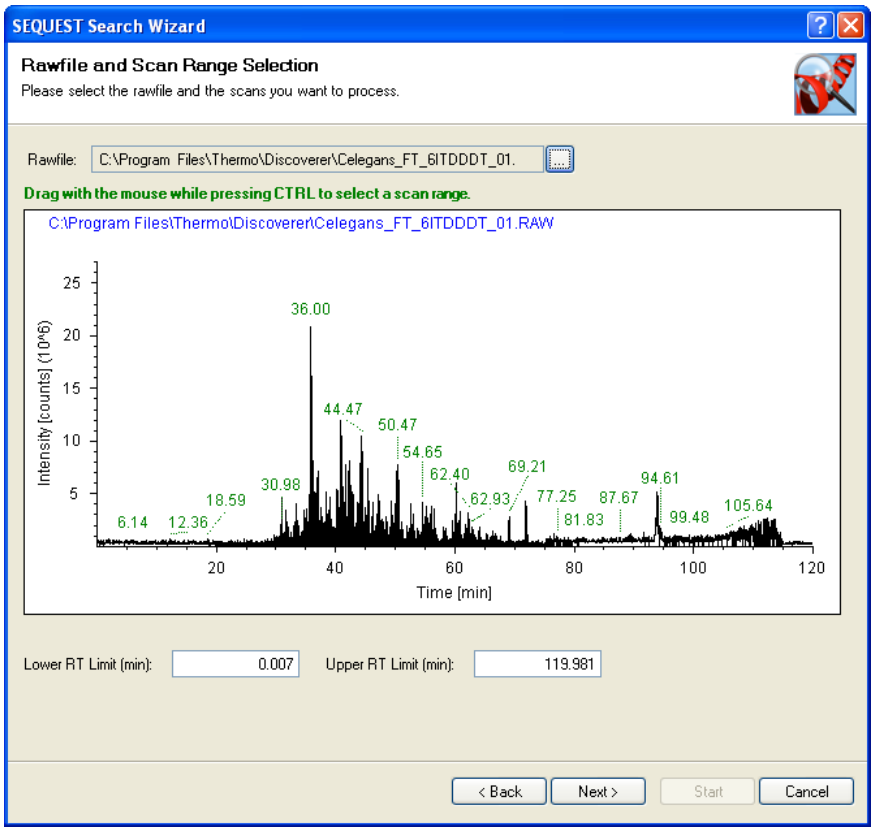
5. On the Rawfile and Scan Range Selection page of the wizard, shown in Figure 40, set your basic search parameters:

   a. In the Rawfile box, click the browse button (...) to search for the raw file in the Open Analysis File(s) dialog box.

   Note The Workflow Editor can accept multiple input raw data files, but the search wizards cannot. For information about creating a workflow for multiple input raw data files, see "Starting a New Search by Using the Workflow Editor" on page 89.

   A base peak chromatogram for the raw data file appears on the page, as shown in Figure 40.

Figure 40. Rawfile and Scan Range Selection page



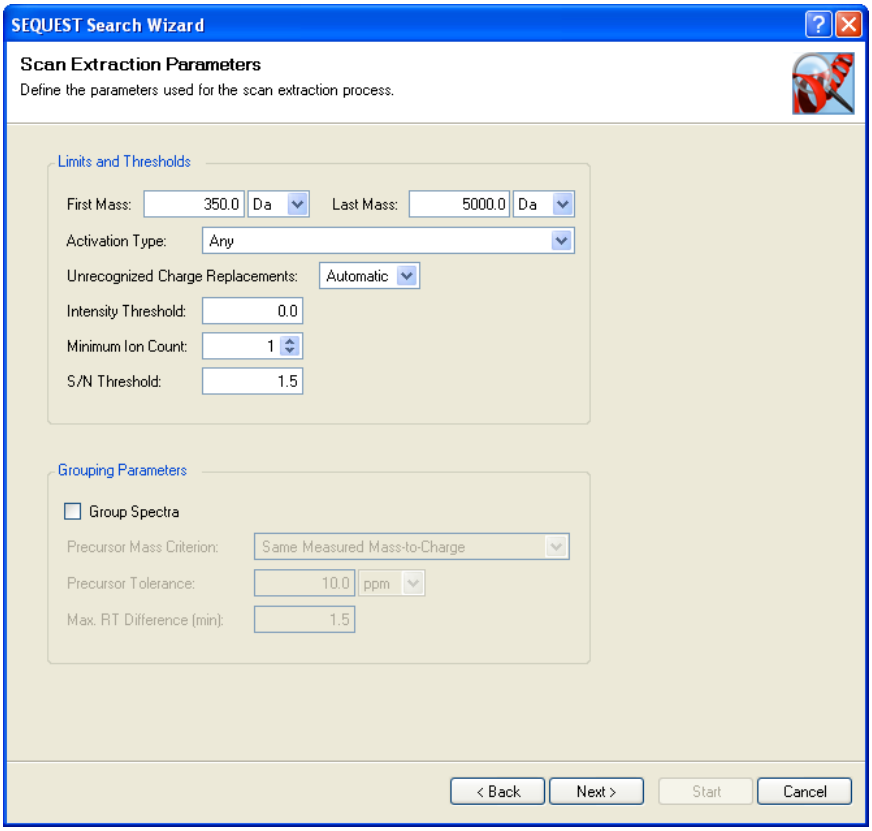
b. Select the range of data to use by choosing either of these methods:

- Hold down the CTRL key and drag the cursor over the range.
- Enter the beginning of the range in the Lower RT Limit (min) box. Enter the end of the range in the Upper RT Limit (min) box.

You might want to exclude the first few minutes of collected data in the raw data file because they contain no peptides or exclude the last few minutes because of cleanup at the end of the data collection.

6. Click **Next**.
7. On the Scan Extraction Parameters page, shown in Figure 41, set the scan extraction parameters.

Figure 41. Scan Extraction Parameters page



a. In the First Mass box, type the mass of the first precursor ion, in daltons. In the Last Mass box, type the mass of the last precursor ion, in daltons.

   These two parameters define the range of ion fragments to search for in the database.

b. From the Activation Type list, select the fragmentation method to use to activate the scan:

   - CID (Collision-Induced Dissociation)
   - MPD (Multi-Photon Dissociation)
   - ECD (Electron Capture Dissociation)
   - PQD (Pulsed Q Collision-Induced Dissociation)
   - ETD (Electron Transfer Dissociation)
   - HCD (High-Energy Collision Dissociation)
   - Any Activation Type

   See "Fragmentation Methods" on page 8 for descriptions of these methods.

   The default is Any Activation Type.

c. In the Unrecognized Charge Replacements list, select the charge number of the precursor ions.

   From the data in the raw file, the Proteome Discoverer application evaluates the spectrum and uses an algorithm to determine the charge state of the spectrum. It cannot calculate the mass without knowing the charge state of the spectrum. If the algorithm cannot determine the charge state of the evaluated spectrum, the application assigns the charge state that you select to the spectrum. You can assign the following charge number:

   - Automatic: Assigns a charge number of +2 and +3 to the spectrum.
   - 1 through 8: Assigns a charge number of from 1 through 8 to the spectrum.

   The default is Automatic.

d. In the Intensity Threshold box, enter an intensity value below which to filter out ions.

   The Proteome Discoverer application filters out low-intensity ions, which are ions that are most likely chemical noise and serve only to slow down the analysis without improving the results.

   The default is 0.0.

e. In the Minimum Ion Count box, enter a value for the minimum ion count or use the increment or decrement buttons.

   The minimum ion count is the minimum number of ions that must be present in an MS/MS spectrum for it to be included in a search.

   The default is 1.

f. In the S/N Threshold box, enter a value for the signal-to-noise threshold setting.

   This setting specifies the intensity of the signal to the intensity of the background noise. It filters out low-intensity ions that function as noise.

   The default is 3.0.

g. (Optional) Select the **Group Spectra** check box.

   The rest of the boxes in the Grouping Parameters area become available.

   In the Grouping Parameters area, you can set grouping parameters to group similar spectra in the raw data file into a single spectrum.

   Grouping spectra speeds up the analysis. The application evaluates an ion only once rather than every time it is observed within the given retention-time limits.

h. In the Precursor Mass Criterion list, select the criteria for grouping. You can select either of these settings:

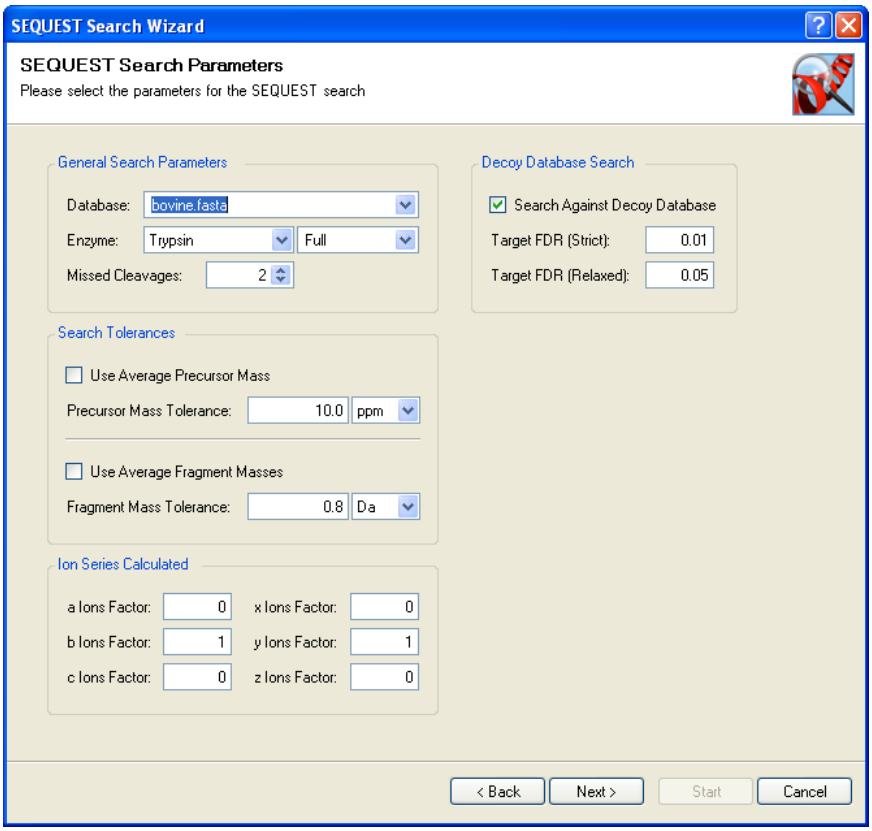
- Same Measured Mass-to-Charge: Groups spectra according to the mass-to-charge ratio (*m/z*) of the precursor ion.
- Same Singly Charged Mass: Groups all charge states with the same singly charged precursor mass. For example, this option groups +2 and +3 ions for the same peptide because they have the same singly charged parent.

i. In the Precursor Tolerance box, type the range of the precursor tolerance, in daltons (Da), milli-mass units (mmu), or parts per million (ppm). For example, if the mass-to-charge ratio of a spectrum is 100.0001 Da and the tolerance is 2 Da, all the spectra with masses in the range of 100.0001 plus or minus 2 Da are valid mass candidates.

j. In the Max. RT Difference (min) box, enter the maximum retention time, in minutes. Retention time is the time in the mass chromatogram when any particular precursor ion is observed. This parameter limits the maximum retention-time difference between scans to be considered for grouping. In general, if the precursor masses of spectra are within the tolerance and the maximum retention time window, they are grouped into a single spectrum. The default is 1.5.

8. Click **Next**.
9. On the SEQUEST Search Parameters page, shown in Figure 42, set the SEQUEST search parameters.

Figure 42. SEQUEST Search Parameters page



a. In the Database list in the General Search Parameters area, select one of the FASTA databases that you have registered.

b. In the Enzyme list, select the enzyme used for digestion and indicate whether the cleavage is full or partial.

   The default enzyme is trypsin, and the default cleavage is Full.

c. In the Missed Cleavages box, use the increment and decrement buttons to specify the maximum number of internal cleavage sites per peptide fragment that is acceptable for an enzyme to miss when cleaving peptides during digestion.

   Normally, the digestion time is too short to enable the enzyme to cleave the protein at all allowed positions, so you must specify the number of missed positions in one resulting peptide fragment where the enzyme could cleave but did not. The minimum value is 0, and the maximum value is 12. The default is 2.

Note The following parameters are also available in the General Search parameters in Mascot:

- Instrument: Specifies the instrument used to process the data in the raw data file.
- Taxonomy: Specifies the category of organism in the Linnaean biological classification system from which the sample was drawn.

In the Search Tolerances area, specify the precursor mass search tolerance.

d. Select the **Use Average Precursor Mass** option to use the average mass for matching the precursor.

e. In the Precursor Mass Tolerance box, specify the precursor mass tolerance value used for finding peptide candidates, in daltons (Da), milli-mass units (mmu), or parts per million (ppm).

- For daltons, the minimum value is 0.0001 and the maximum value is 5.0.
- For milli-mass units, the minimum value is 0.1, and the maximum value is 5000.
- For parts per million, the minimum value is 0.01, and the maximum value is 5000. The default is 10.0.

In the Search Tolerances area, specify the fragment mass search tolerance.

f. Select the **Use Average Fragment Masses** option to use the average mass for matching the fragments.

g. In the Fragment Mass Tolerance box, specify the mass tolerance value used for matching fragment peaks, in daltons (Da) or milli-mass units (mmu).

- For daltons, the minimum value is 0.0001, and the maximum is 2.0. The default is 0.8.
- For milli-mass units, the minimum value is 0.1, and the maximum value is 2000.

h. In the Ion Series Calculated area, specify the ion factors for a, b, c, x, y, and z ions for your experiment type.

You can use a range of 0 through 1.0 for all ion factors. For CID, HCD, and PQD activation types, use b and y ion factors. For ETD and ECD activation types, use c, y, and z ion factors.

Note The Ion Series Calculated area does not appear in the Mascot and ZCore search wizards.

i. (Optional) Set up a decoy database by selecting the **Search Against Decoy Database** check box and setting the false discovery rate (FDR) parameters. For detailed information about this procedure, see "Calculating False Discovery Rates" on page 415.

A decoy database gives a probability value to identifiers and the percentage of false discoveries that you can expect, typically 1 percent.

Note You must select the Search Against Decoy Database check box to see peptide confidence determined by FDR.

- To specify a strict target false discovery rate for peptide matches with high confidence, type a value of **0.0** through **1.0** in the Target FDR (Strict) box.

  The default is 0.01 (1 percent FDR).

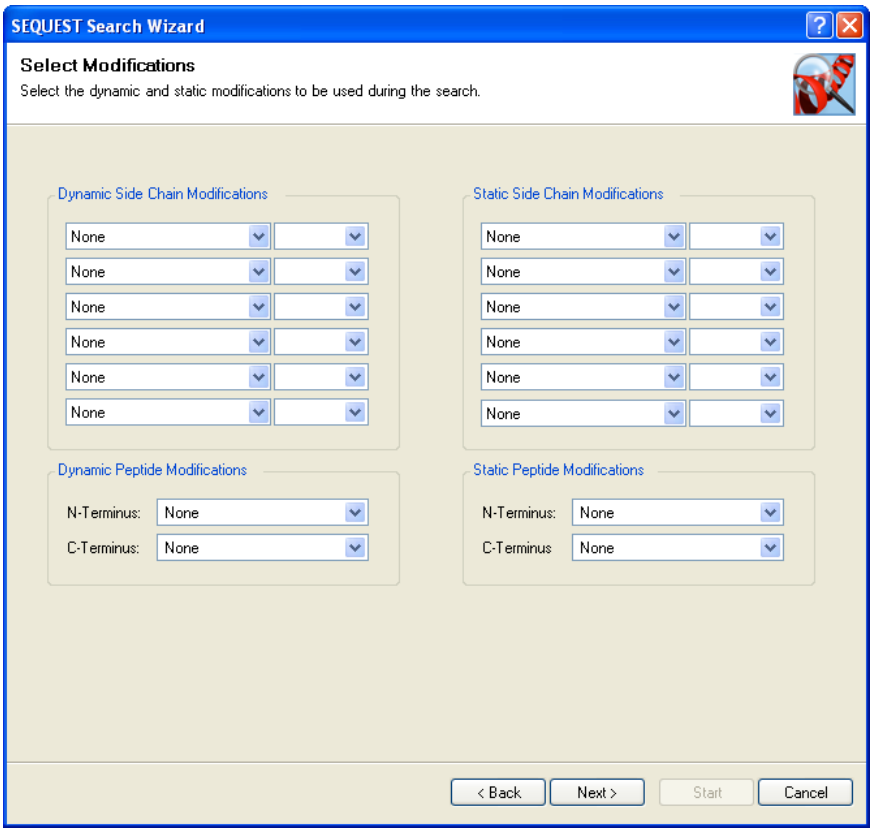
- To specify a relaxed target false discovery rate for peptide matches with moderate confidence, type a value of **0.0** through **1.0** in the Target FDR (Relaxed) box. The default is 0.05 (5 percent FDR).

j. Click **Next**.

10. On the Select Modifications page, shown in Figure 43, specify which modifications you want the search algorithm to include during its in-silico digestion of the protein database.

    For a description of static and dynamic modifications, see "Chemical Modifications" on page 195.

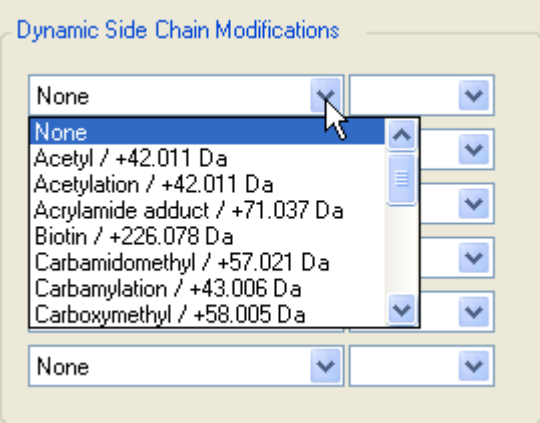
Figure 43. Select Modifications page



a. If you are searching for dynamic modifications, select the modifications and the amino acids on which they can occur in the Dynamic Side Chain Modifications area. In the boxes on the left, select the modifications. In the boxes on the right, select the amino acids on which the modifications occur.

In the SEQUEST and ZCore wizards, delta masses appear next to the names of the modifications in the modification lists to clearly identify the modification, as shown in Figure 44.

Figure 44. Modifications with identifying delta masses



Note In the Mascot wizard, the Dynamic Modifications area replaces both the Dynamic Side Chain Modifications and Dynamic Peptide Modifications areas. You set these modifications on the Mascot server.

The Mascot wizard does not identify by delta masses the modifications that appear on the modification lists as do the SEQUEST and ZCore wizards.

b. If you are searching for static modifications, select the modifications and the amino acids on which they can occur in the Static Side Chain Modifications area. In the boxes on the left, select the modifications. In the boxes on the right, select the amino acids on which the modifications occur.

Note In the Mascot wizard, the Static Modifications area replaces both the Static Side Chain Modifications and Static Peptide Modifications areas. You set these modifications on the Mascot server.

The modifications that appear on the modification lists in the Mascot wizard are not identified by delta masses as they are in the SEQUEST and ZCore wizards.

c. In the N-Terminus list in the Dynamic Peptide Modifications area, select the dynamic modification that occurs on the N terminus of the peptide.

d. In the C-Terminus list in the Dynamic Peptide Modifications area, select the dynamic modification that occurs on the C terminus of the peptide.

Note The Dynamic Peptide Modifications area does not appear in the ZCore wizard.

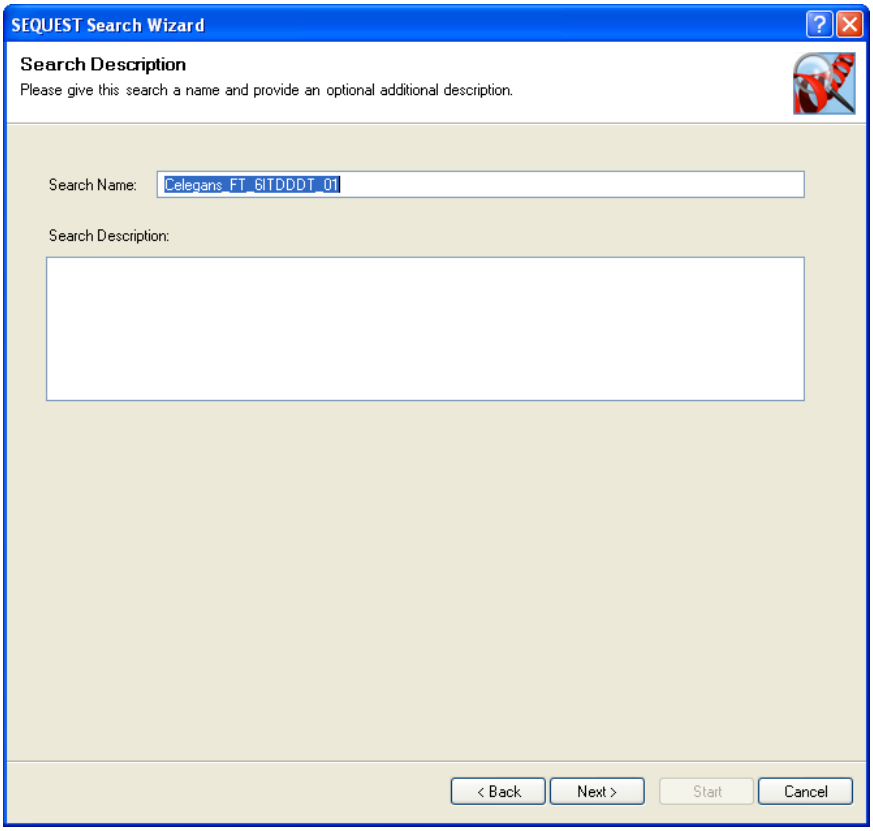
e. In the N-Terminus list in the Static Peptide Modifications area, select the static modification that occurs on the N terminus of the peptide.

f. In the C-Terminus list in the Static Peptide Modifications area, select the static modification that occurs on the C terminus of the peptide.

g. Click **Next**.

11. On the Search Description page, shown in Figure 45, give your search a name and a brief description.

Figure 45. Search Description page



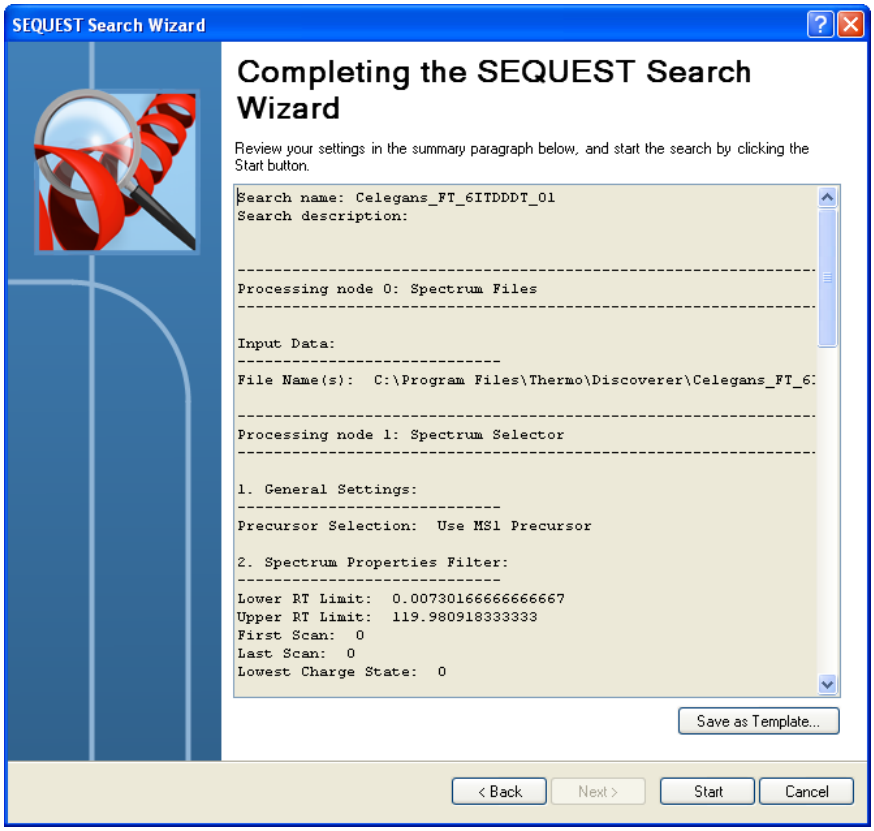
a. In the Search Name box, type a name for your search.

b. In the Search Description box, type a brief description of the search.

c. Click **Next**.

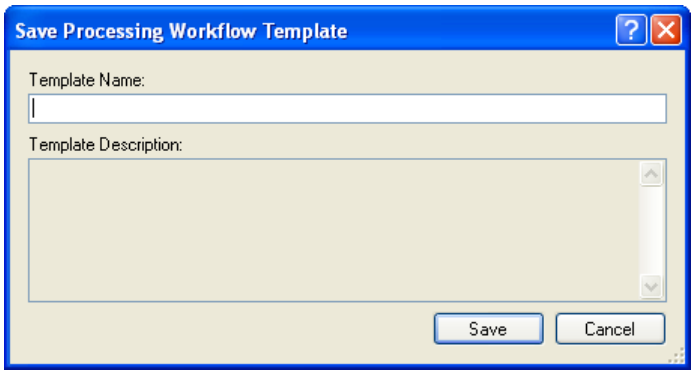
12. (Optional) On the Completing the *Wizard_name* Search Wizard page, shown in Figure 46, click **Save as Template** to save the search parameters that you have set for future use.

Figure 46. Completing the *Wizard_name* Search Wizard page



The Save Processing Workflow Template dialog box appears, as shown in Figure 47.

Figure 47. Save Processing Workflow Template dialog box



a. In the Template Name box, give the search workflow a name.

   The Template Description box reflects the description that you entered on the Search Description page, shown in Figure 45.

b. Click **Save**.

13. Click **Finish** on the Completing the *Wizard_name* Search Wizard page to start the search.

    You can monitor the progress of the search in the job queue. See "Job Queue" on page 28.

14. Choose **File > Open Report** to display your search results. See "Interpreting Search Results" on page 213.
    a. Filter and sort your results. See "Filtering Data" on page 379.
    b. Use different views to aid in your analysis. See "Using the Results Report" on page 213.

Search Wizard Parameters

The following sections describe the parameters on the pages of the SEQUEST Search Wizard.

- Welcome to the SEQUEST Search Wizard Page
- Rawfile and Scan Range Selection Page
- Scan Extraction Parameters Page
- SEQUEST Search Parameters Page
- Modifications Page
- Search Description Page
- Completing the Wizard_name Search Wizard Page

Welcome to the SEQUEST Search Wizard Page

Table 13 describes the parameters on the Welcome to the SEQUEST Search Wizard page.

Table 13. Welcome to the SEQUEST Search Wizard page parameters

Parameters	Description
Templates	Displays a list of the templates that you saved with the Save As Template command on the Completing the SEQUEST Search Wizard page of the search wizard.
Rename	Opens the Renaming Template dialog box so that you can rename the template selected in the Templates box.
Delete	Deletes the template selected in the Templates box.

Rawfile and Scan Range Selection Page

Table 14 describes the parameters on the Rawfile and Scan Range Selection page of the SEQUEST search wizard.

Table 14. Rawfile and Scan Range Selection page parameters

Parameters	Description
Rawfile	Specifies the path and name of the raw file.
Base peak ion chromatogram	Displays a base peak ion chromatogram. • Header: Displays the path and name of the raw file. • Intensity [counts]: Displays the intensity of the largest peak in the spectrum. • Time [min]: Displays the retention time.
Lower RT Limit (min)	Specifies the beginning of the retention time (scan range).
Upper RT Limit (min)	Specifies the end of the retention time (scan range).

Scan Extraction Parameters Page

Table 15 describes the parameters on the Scan Extraction Parameters page of the SEQUEST search wizard.

Table 15. Scan Extraction Parameters page parameters (Sheet 1 of 3)

Parameters	Description
Limits and Thresholds	Sets the parameters for the scan extraction process.
First Mass	Specifies the mass of the first precursor ion in the range of ion fragments to search for in the database, in daltons.
Last Mass	Specifies the mass of the last precursor ion in the range of ion fragments to search for in the database, in daltons.

Table 15. Scan Extraction Parameters page parameters (Sheet 2 of 3)

Parameters	Description
Activation Type	Specifies the fragmentation method to use to activate the scan: • CID (Collision-Induced Dissociation) • MPD (Multi-Photon Dissociation) • ECD (Electron Capture Dissociation) • PQD (Pulsed Q Collision-Induced Dissociation) • ETD (Electron Transfer Dissociation) • HCD (High-Energy Collision Dissociation) • Any Activation Type (Default) For descriptions of these methods, see "Fragmentation Methods" on page 8.
Unrecognized Charge Replacements	Specifies the charge number of the precursor ions. • Automatic (Default): Assigns a charge number of +2 and +3 to the spectrum. • 1 through 8: Assigns a charge number of from 1 through 8 to the spectrum.
Intensity Threshold	Specifies the intensity threshold below which the application filters out ions. The default is 0.0.
Minimum Ion Count	Specifies the minimum ion count, which is the minimum number of ions that must be present in an MS/MS spectrum for it to be included in a search. The default is 1.
S/N Threshold	Specifies the signal-to-noise threshold setting, which is the intensity of the signal to the intensity of the background noise. This threshold filters out low-intensity ions that function as noise. The default is 3.0.
Grouping Parameters	Specifies the parameters for grouping spectra into a single spectrum.
Group Spectra	Determines whether similar spectra in the raw file are grouped into a single spectrum.

Table 15. Scan Extraction Parameters page parameters (Sheet 3 of 3)

Parameters	Description
Precursor Mass Criterion	Selects the criteria for grouping. You can select either of the following settings: • Same Measured Mass-to-Charge (Default): Groups spectra according to the mass-to-charge ratio of the precursor ion. • Same Singly Charged Mass: Groups all charge states with the same singly charged precursor mass. For example, this option groups +2 and +3 ions for the same peptide, because they have the same singly charged parent.
Precursor Tolerance	Specifies the range of the precursor tolerance, in daltons (Da), milli-mass units (mmu), or parts per million (ppm). The default is 10.0 ppm.
Max. RT Difference	Specifies the maximum retention time, in minutes. The default is 1.5.

SEQUEST Search Parameters Page

Table 16 describes the parameters on the SEQUEST Search Parameters page.

Table 16. SEQUEST Search Parameters page parameters (Sheet 1 of 2)

Parameters	Description
General Search Parameters	Specifies the database, enzyme, and number of missed cleavages for the search.
Database	Selects one of the FASTA databases that you have registered.
Enzyme	Specifies the enzyme used for digestion and indicates whether the cleavage is full or partial. The default enzyme is trypsin, and the default cleavage is Full.
Missed Cleavages	Specifies the maximum number of internal cleavage sites per peptide fragment that is acceptable for an enzyme to miss when cleaving peptides during digestion. The minimum value is 0, and the maximum value is 12. The default is 2.
Search Tolerances	Specifies the types of tolerances to use in the search.
Use Average Precursor Mass	Determines whether the average mass is used for matching the precursor.

Table 16. SEQUEST Search Parameters page parameters (Sheet 2 of 2)

Parameters	Description
Precursor Mass Tolerance	Specifies the precursor mass tolerance value used for finding peptide candidates, in daltons (Da), milli-mass units (mmu), or parts per million (ppm). • For daltons, the minimum value is 0.000, and the maximum value is 5.0. • For milli-mass units, the minimum value is 0.1, and the maximum value is 5000. • For parts per million, the minimum value is 0.01, and the maximum value is 5000. The default is 10.0.
Use Average Fragment Masses	Determines whether the average mass is used for matching the fragments.
Fragment Mass Tolerance	Specifies the mass tolerance value used for matching fragment peaks, in daltons (Da) or milli-mass units (mmu). • For daltons, the minimum value is 0.0001, and the maximum value is 2.0. The default is 0.8. • For milli-mass units, the minimum value is 0.1, and the maximum value is 2000.
Ion Series Calculated	Specifies the ion factors for a, b, c, x, y, and z ions for the experiment type. You can use a range of 0 through 1.0 for all ion factors.
Decoy Database Search	Sets the parameters for a decoy database search.
Search Against Decoy Database	Determines whether the Proteome Discoverer application uses a decoy database in the search.
Target FDR (Strict)	Specifies a strict target false discovery rate for peptide matches with high confidence. Use a value of 0.0 through 1.0. The default is 0.01 (1% FDR).
Target FDR (Relaxed)	Specifies a relaxed target false discovery rate for peptide matches with moderate confidence. Use a value of 0.0 through 1.0. The default is 0.05 (5% FDR).

Modifications Page

Table 17 describes the parameters on the Modifications page.

Table 17. Modifications page parameters

Parameters	Description
Dynamic Side Chain Modifications	Specifies the dynamic side-chain modifications on the left and the amino acids where the modifications occur on the right.
Dynamic Peptide Modifications	Specifies the dynamic modifications that you want the search algorithm to include during the in-silico digestion of the protein database.
N-Terminus	Specifies the dynamic modification that occurs on the N terminus of the peptide.
C-Terminus	Specifies the dynamic modification that occurs on the C terminus of the peptide.
Static Side Chain Modifications	Specifies the static side-chain modifications on the left and the amino acids where the modifications occur on the right.
Static Peptide Modifications	Specifies the static modifications that you want the search algorithm to include during the in-silico digestion of the protein database.
N-Terminus	Specifies the static modification that occurs on the N terminus of the peptide.
C-Terminus	Specifies the static modification that occurs on the C terminus of the peptide.

Search Description Page

Table 18 describes the parameters on the Search Description page.

Table 18. Search Description page parameters

Parameters	Description
Search Name	Specifies the name of the search.
Search Description	Briefly describes the search.

Completing the *Wizard_name* Search Wizard Page

Table 19 describes the parameters on the Completing the *Wizard_name* Search Wizard page.

Table 19. Completing the *Wizard_name* Search Wizard page parameters

Parameters	Description
Save As Template	Opens the Save Processing Workflow Template dialog box, shown in Figure 47, so that you can save the search workflow as a template.
Finish	Begins executing the search.

Starting a New Search by Using the Workflow Editor

You can create a customized search by using the Proteome Discoverer Workflow Editor instead of the search wizards. The Workflow Editor is a flexible and complex tool that you can use to create customized data-processing workflows. Instead of using the standard wizards available through the Processing menu, you can develop a workflow specific to your needs. The Workflow Editor searches with multiple algorithms and merges results from multiple fragmentation methods. It also provides great flexibility in creating custom search results. Unlike the search wizards, the Workflow Editor can accept multiple input raw files.

You can create a reusable processing workflow template by saving your design to load and use at another time. A unique workflow gives you the ability to set parameters that are normally static settings in the wizard or use a function that would not normally be available, such as deconvoluting the precursor ions for all high-mass-accuracy data or exporting a spectrum.

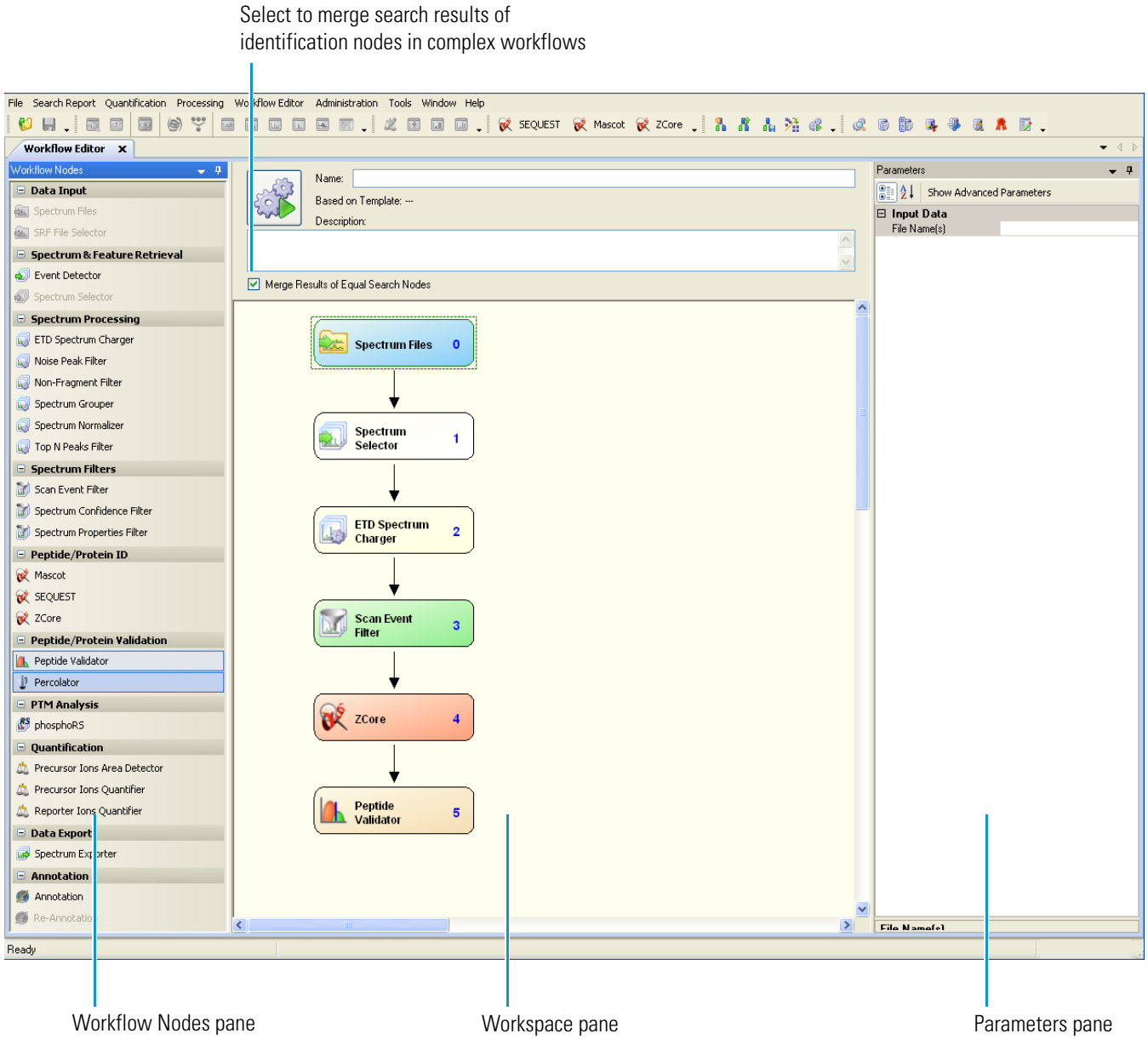
The workflow is the layout of processing nodes, or workflow steps, which you then submit to process your data. The nodes are like building blocks that you can use to create a unique search sequence. You can use them to define your own search parameter tolerances and criteria.

WARNING As a prerequisite to using the Proteome Discoverer Workflow Editor, you must know how each workflow node functions. If you do not understand the function (or interconnectivity) of these nodes, you can potentially build a sequence that creates bad results and makes no analytical sense. For a brief description of these nodes, see "Workflow Nodes" on page 117. For a more detailed description, see "Workflow Editor Nodes" on page 637.

You can access the Workflow Editor through the Workflow Editor menu in the Proteome Discoverer application or through the Workflow Editor icons on the main toolbar. After you choose a menu command or click an icon, the application opens a Workflow Editor page in the main window.

The three-pane layout of the Workflow Editor page provides a pane for node selections, a workspace for placing the nodes, and a pane where you can choose parameters for each node, as shown in Figure 48.

Figure 48. Workflow Editor workspace

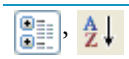


To create a workflow, see "Creating a Search Workflow" on page 91.

Workflow Editor Parameters

Table 20 describes the parameters available in the Workflow Editor.

Table 20. Workflow Editor page parameters

Workflow node	Function
Workflow Nodes	Lists the nodes available for constructing a workflow. For a detailed description of each node shown in this pane, see "Workflow Editor Nodes" on page 637.
Name	Specifies the name of the workflow.
Based on Template	Displays the name of any existing template that you choose when you import an existing workflow.
Description	Displays a description of the workflow.
Merge Results of Equal Search Nodes	Determines whether peptides and proteins identified by the same search engine type are merged together, using a single search algorithm. For example, you can only merge SEQUEST search results with SEQUEST search results of a different node. This option is selected by default.
Workspace pane	Contains the selected nodes connected together in a workflow.
Parameters	Sets the search parameters for each node in the workspace.
	For information about these icons, see "Dialog Box Icons" on page 35.
Show/Hide Advanced Parameters	Determines whether more advanced parameters are displayed or hidden in the Parameters pane.

Creating a Search Workflow

You can use the following procedure to process one raw file from one sample, multiple raw files from one sample, or multiple raw files from multiple samples. For additional details on creating a workflow for multiple raw files from one sample, see "Creating a Search Workflow for Multiple Raw Files from the Same Sample" on page 101.

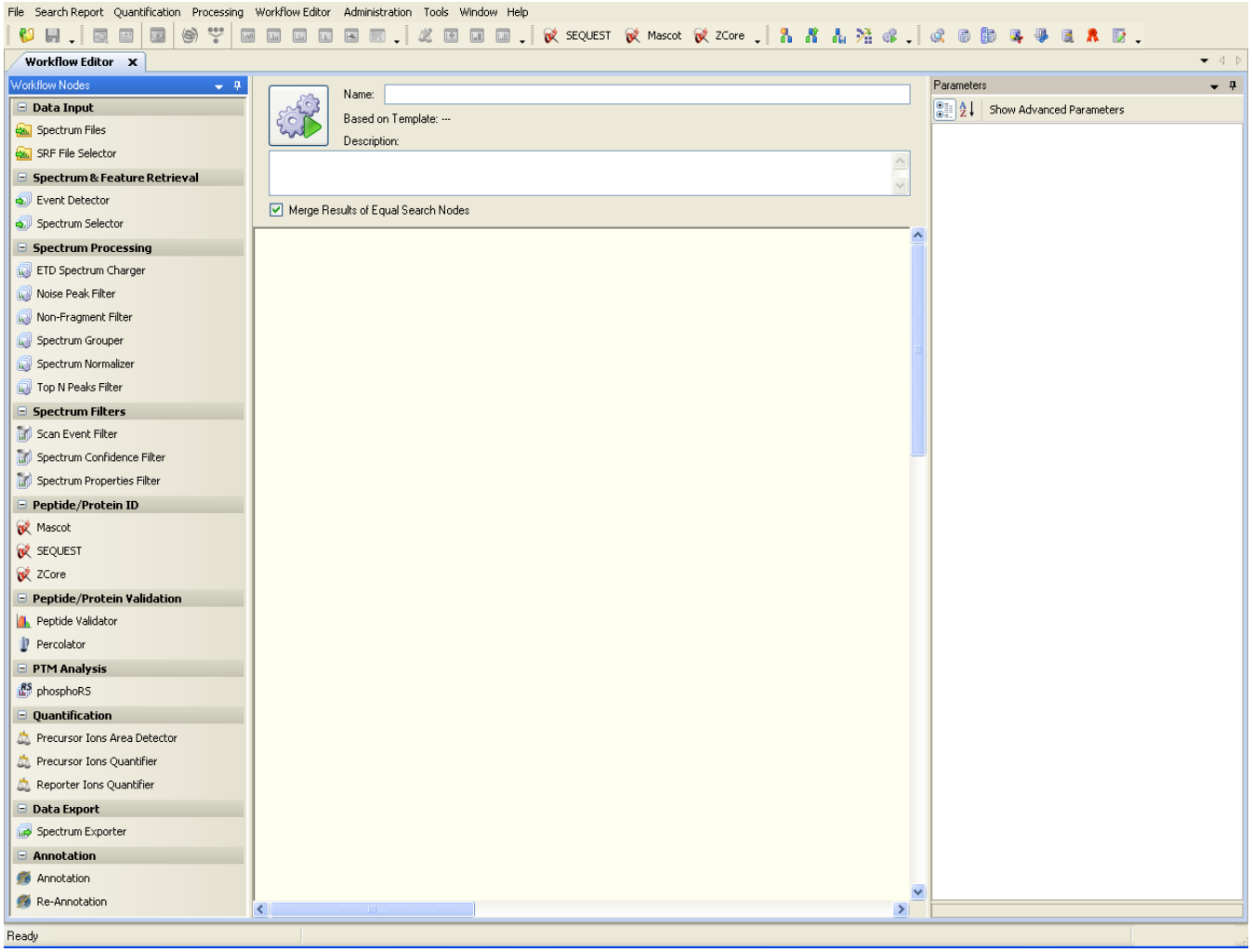
For a demonstration showing how to create a new workflow, see "Demonstrating How to Create a Workflow" on page 99.

❖ To create a new workflow

1. Choose **Workflow Editor > New Workflow** or click the **New Workflow** icon,  .

   The Workflow Editor opens, as shown in Figure 49.

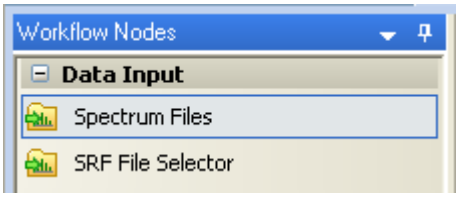
Figure 49. Workflow Editor



2. In the Name box in the workspace pane, type a name for the workflow.
3. (Optional) In the Description box, type a description of the workflow.
4. To perform two searches using the same search engine node and then merge the search results in the output MSF file, select the **Merge Results of Equal Search Nodes** check box.
5. From the Data Input area of the Workflow Nodes pane, drag the **Spectrum Files** or **SRF File Selector** node to the workspace pane.

   Figure 50 shows a selected Data Input node.

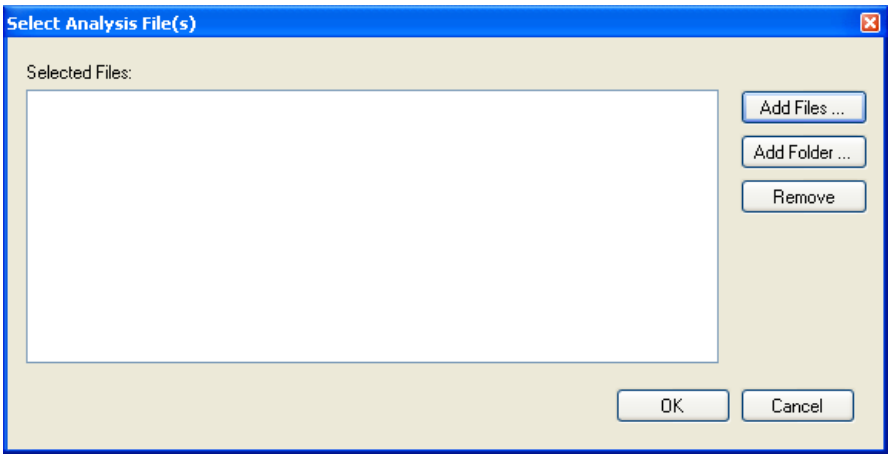
Figure 50. Selected data input node in the Workflow Editor



6. Select the **Spectrum Files** node if it is not already selected.
7. Select the data input file:
   a. In the Input Data section at the top right of the Parameters pane, click the File Name(s) row (see Figure 48 on page 90).
   b. Click the browse button (...) in that row.

      The Select Analysis File(s) dialog box appears, as shown in Figure 51.

      Figure 51. Select Analysis File(s) dialog box



   c. Click **Add Files** to open the Add Analysis File(s) dialog box.
   d. Browse to the location of the data input file, select the file, and click **Open**.
   e. Click **OK** to close the Select Analysis File dialog box.
8. If you selected the Spectrum Files node in step 5, drag the **Spectrum Selector** node to the workspace and place it beneath the Spectrum Files node.

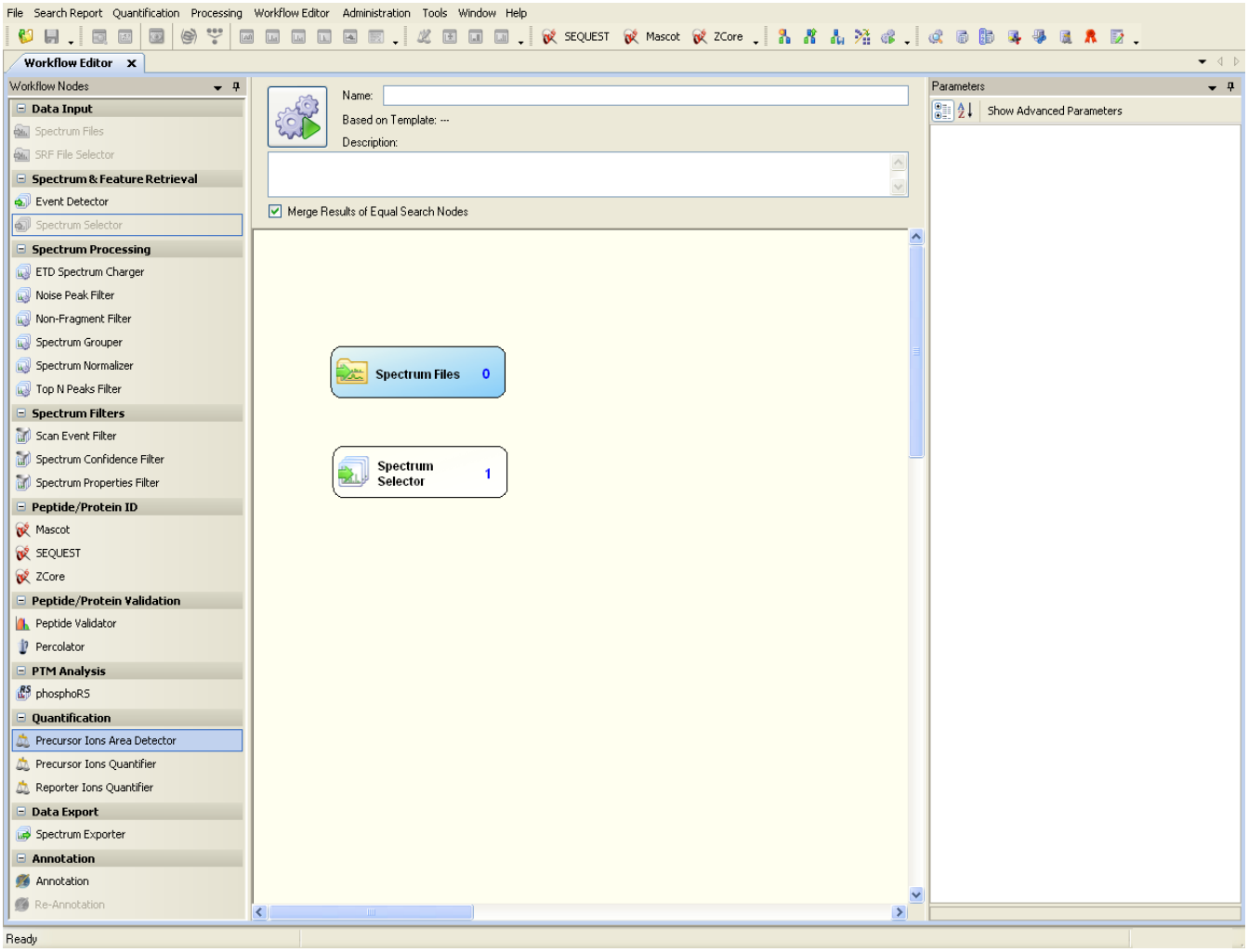
> **Note** If you selected the SRF File Selector node instead of the Spectrum Files node, you do not have to drag the Spectrum Selector node to the workspace pane.

Figure 52 shows the addition of the Spectrum Files and Spectrum Selector nodes to the workspace. Selecting the Spectrum Selector node in the workspace pane displays the available parameters for that node in the right pane.

The numbers that appear on each workflow node indicate the order in which the Proteome Discoverer application processes the nodes.

Note You can set the Spectrum Selector node to select which precursor mass to use for a given MS^n scan, such as choosing the precursor from the parent scan.

Figure 52. Spectrum Files and Spectrum Selector nodes added to a workflow



9. Depending on your data needs, drag the appropriate nodes from the Workflow Nodes pane to the workspace pane.

   For a description of the nodes that you can select, see "Workflow Nodes" on page 117 or "Workflow Editor Nodes" on page 637. The nodes in each section of the Workflow Nodes pane appear in unique colors; for example, the Data Input nodes are blue, the quantification nodes are pink, and the Spectrum Processing nodes are yellow.

   When you use a search engine node like SEQUEST or Mascot in the workflow, you must attached the Peptide Validator or the Percolator node to it.

You can also add third-party nodes that are in your installation that are not documented in this manual. For further information on those nodes, consult the third-party documentation.

You cannot drag workflow nodes into the workspace pane that cannot logically be added at that point. For example, you cannot add the Spectrum Files node and the SRF File Selector node to the workspace together because they are mutually exclusive.

10. Organize the nodes to reflect a procedural order from top to bottom so that the Spectrum Files node remains on top as the root node.

    Delete a node by selecting the node in the workspace pane and pressing DELETE or by right-clicking the node and choosing **Cut** (or **CTRL+X**) from the shortcut menu.

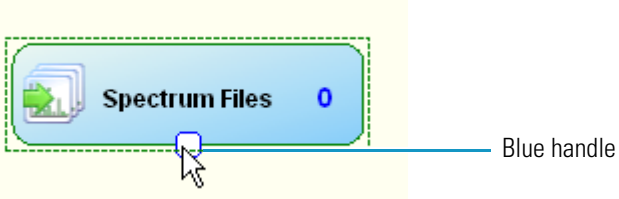
    You can use the Cut command and the Paste (or CTRL+P) command on the shortcut menu to move a node to another place in the workspace or use the Copy (or CTRL+C) and Paste commands to duplicate a node in the workspace.

    You can paste copied or cut nodes into other workflows.

11. Connect the nodes:

    a. Click the top node so that a blue handle is activated at the bottom center of the node, as shown in Figure 53.

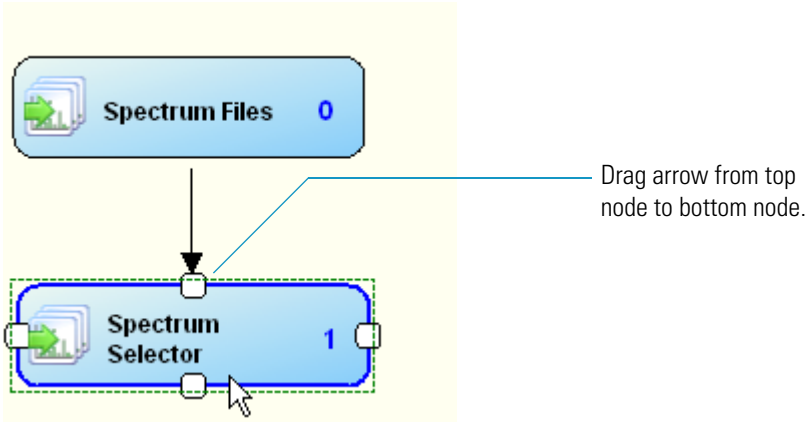
    Figure 53. Activated node example

    Joining the nodes together creates a sequence of steps for the Proteome Discoverer application to follow.

    b. Drag the blue handle down to the top-center of the node below it, as shown in Figure 54.

    Figure 54. Joining two nodes

IMPORTANT If the next node appears with a red edge at this point, you cannot connect to the previous node.

If the Workflow Editor prevents you from connecting two nodes, the workflow is erroneous.

c. Link all the nodes to develop a workflow.

12. After you join all your chosen nodes, align them by choosing **Workflow Editor > Auto Layout**, or clicking the **Auto Layout** icon ( ), or right-clicking a node and choosing **Auto Layout** from the shortcut menu.

13. (Optional) You can renumber the workflow nodes in the workflow in consecutive order by choosing **Workflow Editor > Auto Number**.

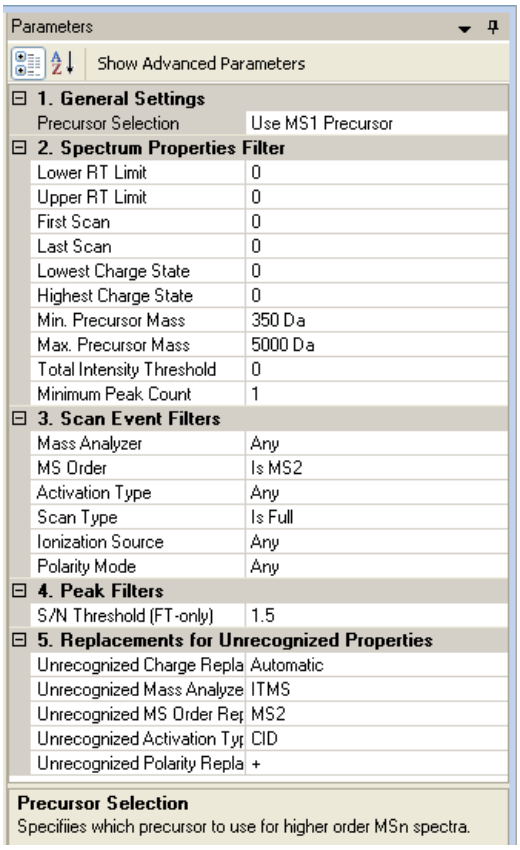
14. Set the parameters for each node in the workspace pane:

a. Click the node to activate its functions.

The available parameters for the node appear in the Parameters pane, as shown in the example for the Spectrum Selector node in Figure 55.

Note The same options are available in the search wizards.

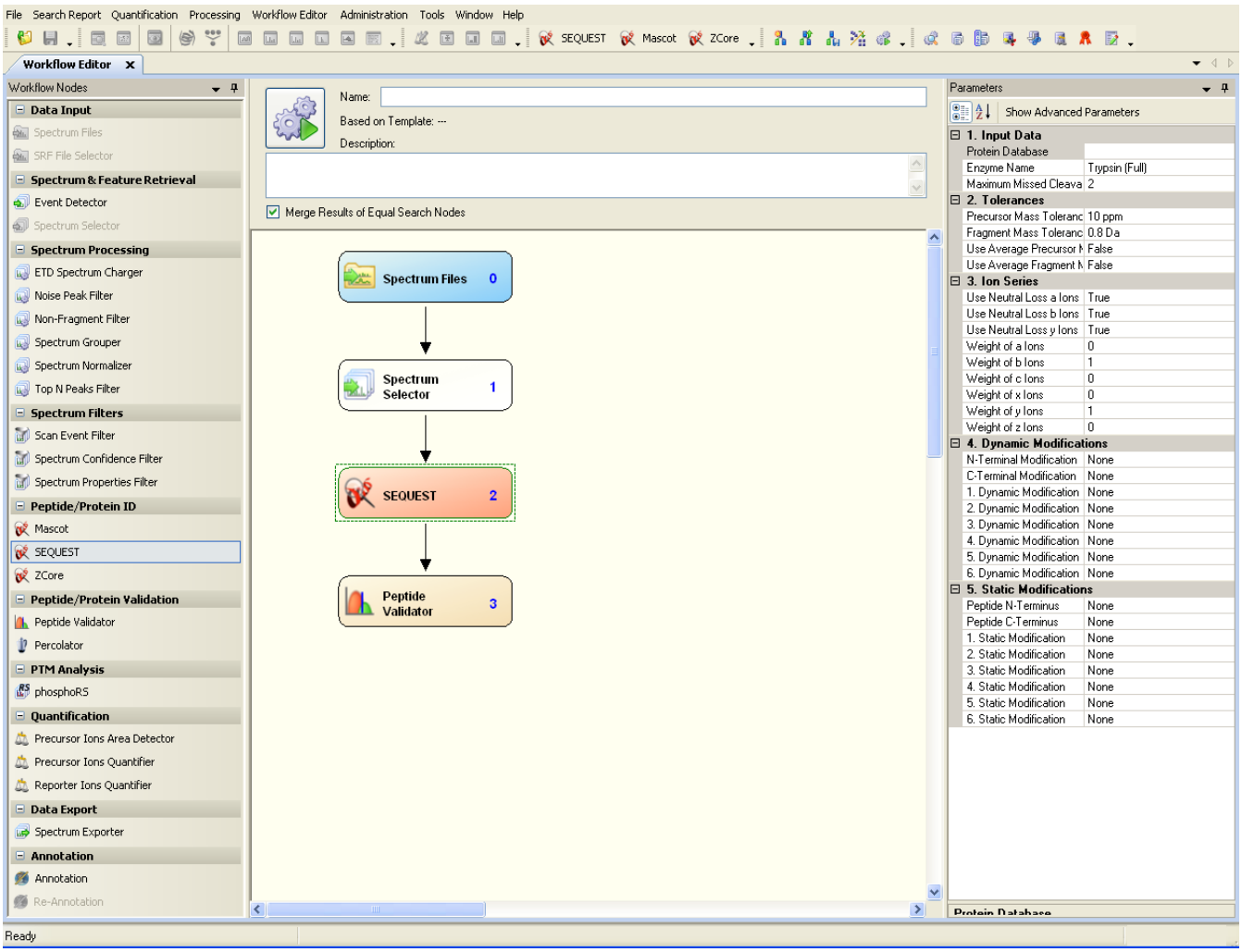
Figure 55. Parameters for Spectrum Selector node in the Parameters pane



b. Set the node's parameters. Complete this step for each node that you select.

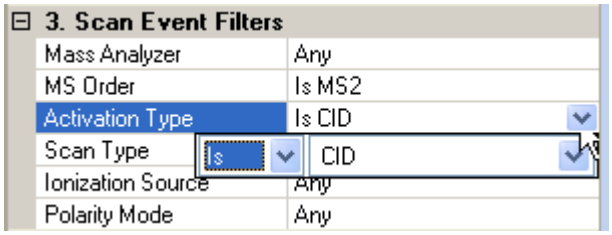
Figure 56 shows the parameters set for the SEQUEST node.

Figure 56. Setting parameters for the workflow



When you click some parameters, two lists appear, as shown in Figure 57.

Figure 57. Settings and filters



You can use the list on the left to apply a filter to the setting that you select in the list on the right. The list on the left consists of three options:

- Is: Applies the setting selected in the list on the right. In the example in Figure 57, "Is" means that the workflow processes data from the CID activation type.
- Is Not: Applies all settings in the list on the right except the selected setting. In the example in Figure 57, "Is Not" means that the workflow processes data from all activation types except CID.
- Any: Applies all settings available for the parameter in the list on the right. In the example in Figure 57, "Any" means that the workflow processes data from any activation type available in the list on the right.

Any is the default.

You can filter input data before searching the database to remove lower-quality spectral peak lists from your analysis. This step might help to decrease search times and false positive identifications. The Spectrum Filters area of the Workflow Nodes pane provides three types of spectrum filters to use for your search. Use these pre-analysis filters to streamline your search results. For information about these nodes, see "Spectrum Filters Nodes" on page 659.

Use the Scan Event Filter node for high-mass-accuracy data, such as Mascot analysis and SEQUEST analysis of mixed fragmentation-mode-type data (CID and ETD). It can filter information according to fragmentation type, mass analyzer identity, and other parameters. See "Scan Event Filter Node" on page 659 for information about the Scan Event Filter node.

❖ To save the workflow as a template

1. Choose **Workflow Editor > Save as Template** or click the **Save As Template** icon,  .

   (To save the workflow in XML format, see "Saving a Workflow as an XML Template" on page 114.)

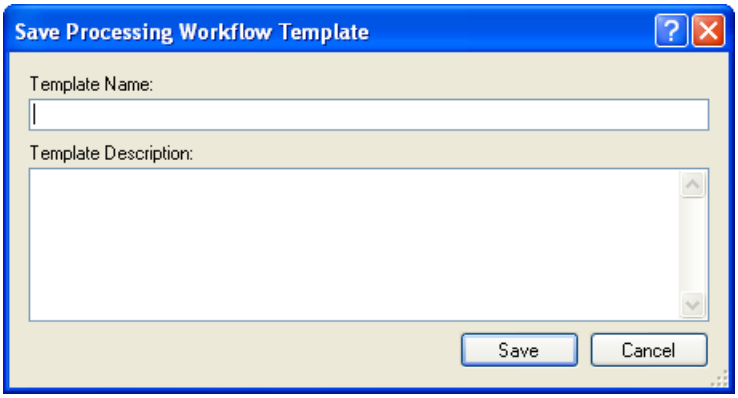
2. In the Save Processing Workflow Template dialog box, shown in Figure 58, do the following:

   a. Type a template name in the Template Name box.

   b. Type a description in the Template Description box.

   c. Click **Save**.

Figure 58. Save Processing Workflow Template dialog box



❖ **To perform your search**

1. Choose **Workflow Editor > Start Workflow** or click the **Start Workflow** icon, .

   The job queue appears, showing the status of your search.

2. Use the job queue to check the status of your search as the search progresses.

   For information about the job queue, see "Job Queue" on page 28.

3. Choose **File > Open Report** to display your search results. See "Opening MSF Files" on page 215.

   a. Filter and sort your results. See "Filtering Data" on page 359.

   b. Use different views to aid in your analysis. See "Using the Results Report" on page 213.

Demonstrating How to Create a Workflow

The following demonstration shows you how to set up a workflow. In this example, a sample containing a trypsin digest of Caenorhabditis elegans, a nematodal worm, was submitted to an LTQ Orbitrap XL mass spectrometer at a resolution of 60 000 for MS/MS processing, using both the ETD and CID fragmentation methods for better confidence. The example searches a FASTA database to determine how the worm's proteins are expressed.

Click here to view the demonstration. To see a full-screen version, right-click on the demonstration after you start it and choose **Full Screen Multimedia**.



Creating a Search Workflow for Multiple Raw Files from the Same Sample

Multidimensional Protein Identification Technology (MudPIT) experiments investigate complex proteomes by applying multidimensional chromatography to the samples before acquisition in the mass spectrometer. Typically, this process results in several dozen or even a few hundred fractions that are separately analyzed by LC-MS, resulting in one raw file per sample fraction. Analyzing gel slices or performing in-depth follow-up acquisitions also results

Creating a Search Workflow for Multiple Raw Files from the Same Sample

Multidimensional Protein Identification Technology (MudPIT) experiments investigate complex proteomes by applying multidimensional chromatography to the samples before acquisition in the mass spectrometer. Typically, this process results in several dozen or even a few hundred fractions that are separately analyzed by LC-MS, resulting in one raw file per sample fraction. Analyzing gel slices or performing in-depth follow-up acquisitions also results in multiple fractions. Because all these fractions belong to the same sample, the Proteome Discoverer application can process all raw files from these fractions as one contiguous input file and generates only one result file.

You have two ways to search for sample fractions:

- Search the sample fractions one at a time and open them in a multiconsensus report.

  This method is appropriate for searching multiple samples. When you open a multiconsensus report from several searches, the Proteome Discoverer application does not calculate a combined protein score, and it orders the proteins by their coverage.

- Search the fractions all at one time in MudPIT.

  To search the fractions of only one sample, use MudPIT. In this mode, the Proteome Discoverer application searches all fractions as one logical sample and creates a single MSF result file. It automatically merges all identified peptides and proteins from all fractions and creates a single combined score for every protein that includes all peptides identified from the different fractions.

  Opening a MudPIT report is faster and consumes less memory than combining separate reports into a multiconsensus report. For example, if the Proteome Discoverer application identifies a protein in every fraction and opens all fractions into a multiconsensus report, a copy of the same protein resides in memory for every fraction that the protein was identified in. It must merge the proteins into an additional protein instance that it displays in the multiconsensus report, slowing performance and consuming memory unnecessarily. However, if you searched the fractions in MudPIT mode, the proteins are already merged from the different fractions, and the Proteome Discoverer application only needs to load the identified merged proteins.

The following procedure describes how to create a workflow for multiple raw files from the same sample. This workflow is basically the same as that given in "Creating a Search Workflow" on page 91, except that you select multiple files to load with the File Name(s) parameter of the Spectrum Files node.

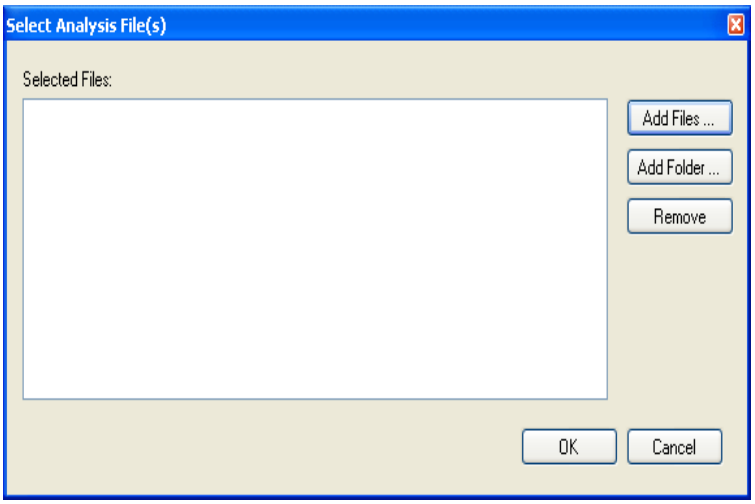
Note The following method is not appropriate for batch-processing different sample data files because the process generates a single result file.

❖ To load multiple raw files from the same sample

1. In the Workflow Editor, drag the **Spectrum Files** node from the Data Input section of the Workflow Nodes pane to the workspace pane.
2. Select the **Spectrum Files** node.
3. In the Parameters pane, click **File Name(s)**, and click the browse button (...).

   The Select Analysis File(s) dialog box appears, as shown in Figure 59.

   Figure 59. Select Analysis File(s) dialog box



4. To add new input files, click **Add Files**, and in the Add Analysis File(s) dialog box, select the raw data files to load and click **Open**.

   –or–

   To add all the raw data files in a specific folder, click **Add Folder**, and in the Browse for File dialog box, click **OK**.

   To remove a file or folder from the Selected Files area of the dialog box, select the file and click **Remove**.

5. In the Select Analysis File(s) dialog box, click **OK**.
6. Drag the Spectrum Selector node to the workspace pane beneath the Spectrum Files node, and continue with the process of creating a workflow, as described in "Creating a Search Workflow" on page 91.
7. Choose **Workflow Editor > Start Workflow** to start the workflow.

You can use the Proteome Discoverer Daemon utility to monitor multiple searches on multiple raw data files. For information about this tool, see "Using the Proteome Discoverer Daemon Utility" on page 121.

Creating a Quantification Workflow

To perform quantification, you must run a quantification workflow. A quantification workflow is a search workflow that includes one of three quantification nodes found in the quantification section of the Workflow Nodes pane of the Workflow Editor. Table 21 lists these nodes and where you can obtain information about creating a quantification workflow for each.

Table 21. Quantification nodes

Quantification node	Use	For more information
Precursor Ions Quantifier node	For precursor ion quantification (for example, SILAC)	See "Performing Precursor Ion Quantification" on page 469.
Reporter Ions Quantifier node	For reporter ion quantification (for example, iTRAQ and TMT)	See "Performing Reporter Ion Quantification" on page 476.
Precursor Ions Area Detector node	For peak area calculation quantification	See "Performing Peak Area Calculation Quantification" on page 483.

You must attach the selected quantification node directly to the Event Detector node. For information about the parameters that you can set for the quantification nodes, see "Quantification Nodes" on page 698.

Creating an Annotation Workflow

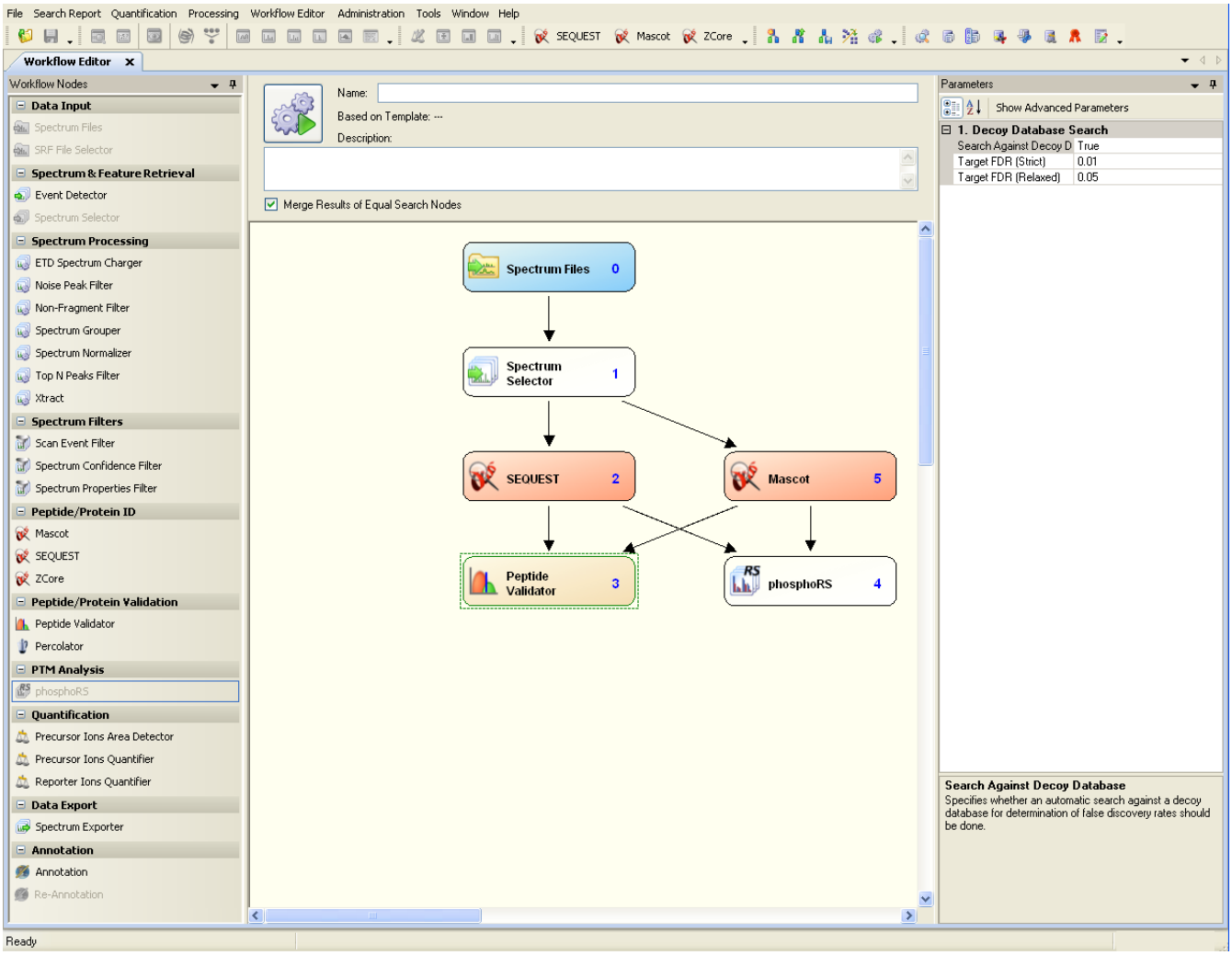
To create a workflow that uses the Annotation node to retrieve GO database information from ProteinCenter and install it in the Proteome Discoverer results files, see "Creating a Protein Annotation Workflow" on page 437.

Creating a PTM Analysis Workflow

If you want to focus on studying the biologically relevant post-translational modifications of proteins, you can create a workflow that includes the phospho*RS* node (see "phosphoRS Node" on page 691). This node calculates PTM site localization scores for phosphorylation and makes them available in the Proteins Identification Details view when you choose Search Report > Protein ID Details View. This view color-codes the found phosphorylation modification above the amino acid sequences to indicate the probability of the modification being found on those portions of the amino acid. The PTM Site Probabilities area to the left of the sequence table displays a legend explaining the color-coding. For more information o this view, see "Interpreting Your Results with the Protein Identification Details View" on page 245.

You can use only one phospho*RS* node in a workflow. Connect it to all search nodes whose results you want to submit to phosphorylation site localization scoring. Figure 60 gives an example of a workflow with two different search nodes attached to the phospho*RS* node.

Figure 60. Workflow with two different search nodes attached to the phospho*RS* node

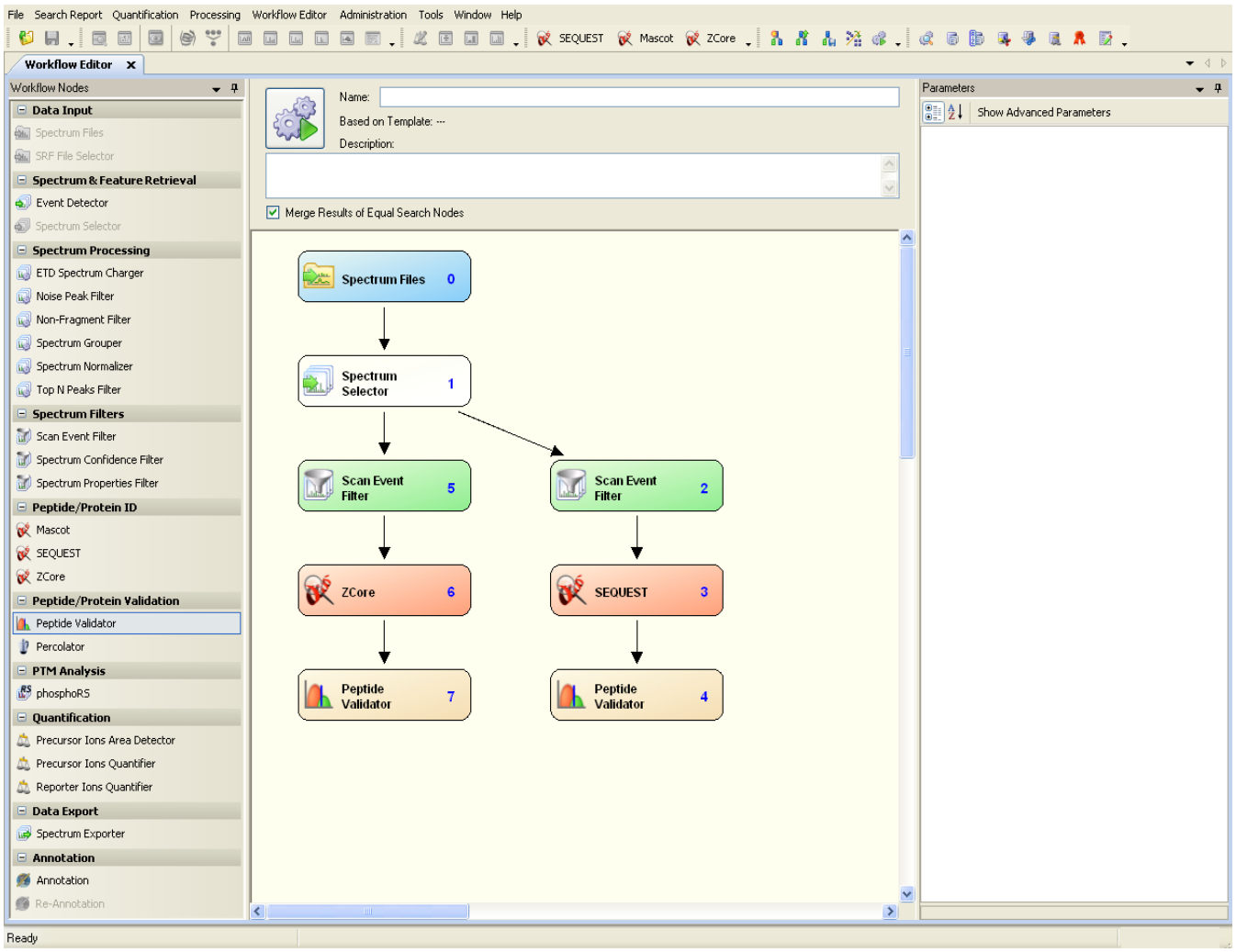


phospho*RS* retrieves the phosphorylation sites that were searched and the mass tolerance used for matching fragment ions directly from the attached search nodes. It has two additional parameters for choosing a specific mass tolerance to use when matching fragment ions (see Table 138). With these parameters, you can overwrite the default mass tolerance setting used in the search node.

Creating Parallel Workflows

Parallel workflows are workflows that search the same raw data file and the same part of the spectrum but specify different criteria, different search nodes for the search, or both. They resemble the example workflow shown in Figure 61. You can use parallel workflows to conduct two or more searches using two or more search engines on the same raw data and to compare the results of these two searches at the same time. For example, you may want to search both CID and ETD data from the same raw data file to increase the chances of finding a match. CID data contains b and y ions, and ETD data contains b, c, and z ions, so the two types of data are complementary. You can also use a parallel workflow for quantification.

Figure 61. Parallel workflow



The following instructions show you how to create the simple parallel workflow shown in Figure 61.

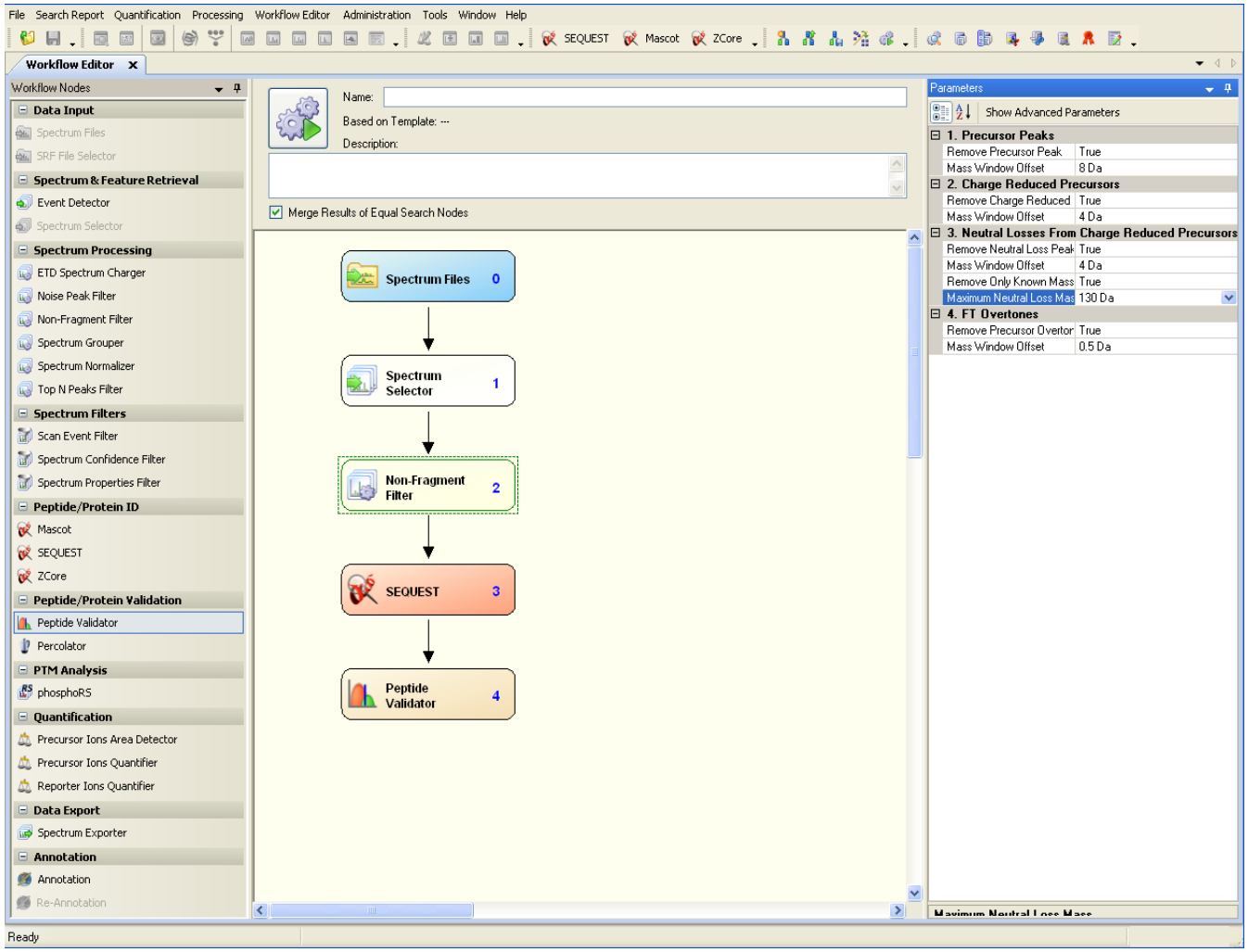
❖ **To create the parallel workflow shown in Figure 61**

1. Drag the Spectrum Files node to the workspace pane, and specify the name and path of the raw data file in the Parameters pane.
2. Drag the Spectrum Selector node to the workspace pane and place it directly under the Spectrum Files node. Set the parameters.
3. Drag two Scan Event Filter nodes to the workspace pane and place them side by side beneath the Spectrum Selector node. In the Parameters pane, set the Activation Type parameter to CID for one node and to ETD for the other node.
4. Drag the SEQUEST node to the workspace pane and place it beneath the Scan Event Filter node set to the CID activation type.
5. Drag the ZCore node to the workspace pane and place it beneath the Scan Event Filter node set to the ETD activation type. (ZCore can only process the ETD data.)
6. Connect the nodes as shown in Figure 61.
7. Choose **Workflow Editor > Start Workflow** to start the parallel workflow.

Adding a Non-Fragment Filter Node for High-Resolution Data

The main purpose of the Non-Fragment Filter node is to remove precursor peaks from the spectra that are not related to peptide fragments and could therefore increase the risk of the search engines making false positive matches. If you add a Non-Fragment Filter node to the workflow for processing data taken from Orbitrap instruments, Thermo Fisher Scientific recommends that you remove most of the precursor peaks. Setting the window to a smaller width increases the risk of leaving some of the precursor peaks or their side bands in the spectrum. Figure 62 shows the recommended settings with wider tolerances.

Figure 62. Non-Fragment Filter node settings for data taken from LTQ Orbitrap instruments



Peaks arising from overtones are rarely seen within Orbitrap spectra but are prominent peaks in spectra from the LT FT instruments. The range in which neutral loss peaks from the charge-reduced precursor peaks are removed is scaled by the charge of the charge-reduced peak. Therefore, if you specify a value of 130 Da, as in Figure 62, the Proteome Discoverer application removes neutral loss peaks within a 130-Da range for +1 peaks, a 65-Da range for +2 peaks, and so forth. To remove neutral losses, you can remove either every peak within the specified range or only those peaks from an internal table of known neutral loss masses from charge-reduced precursor ions, such as those shown in Table 22.

Table 22. Mass of known neutral losses from charge-reduced precursor ions

Mass	Neutral loss
17.027	NH3
18.011	H2O
27.995 Da	CO
32.026 Da	CH3OH
34.053 Da	N2H6 (2xNH3)
35.037 Da	H4NO
36.021 Da	H4O2 (2xH20)
44.037 Da	CH4N2
45.021 Da	CH3NO
46.006 Da	CH202
46.042 Da	C2H6O
59.037 Da	C2H5NO
59.048 Da	CH5N3
73.089 Da	C4H11N
74.019 Da	C3H6S
82.053 Da	C4H6N2
86.072 Da	C3H8N3
99.068 Da	C4H9N3
101.095 Da	C4H11N3
108.58 Da	C7H8O
131.074 Da	C9H9N

Opening an Existing Workflow

You can open an existing workflow from a template that you saved, or you can open it from an MSF or XML file.

Opening an Existing Workflow from a Template

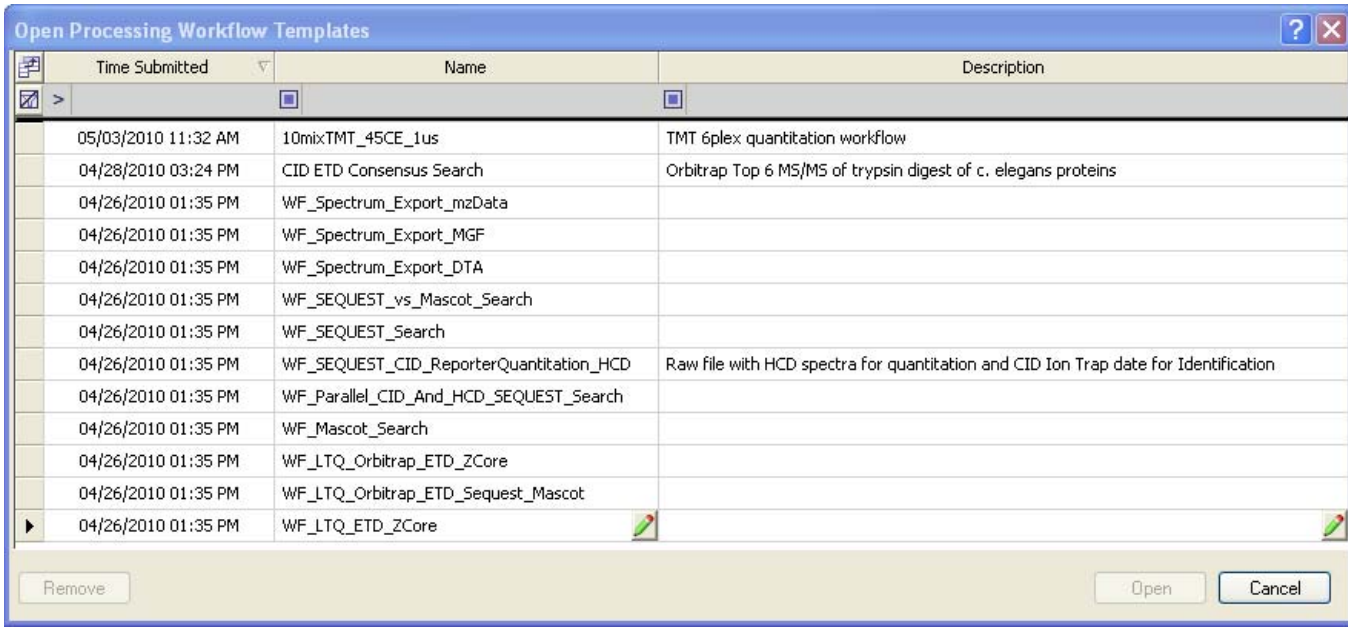
You can open an existing workflow that you previously saved when you chose Workflow Editor > Save As Template.

❖ To open an existing workflow from a template

1. Choose **Workflow Editor > Open From Template** or click the **Open From Template** icon,  .

   The Open Processing Workflow Templates dialog box appears, as shown in the example in Figure 63, listing the available workflow templates.

Figure 63. Open Processing Workflow Templates dialog box

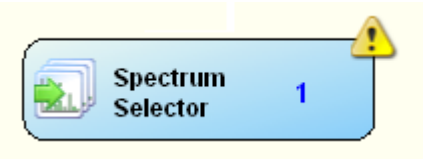


2. Select a workflow from the list.
3. Click **Open**.

The Workflow Editor window opens, displaying the selected workflow. The Based on Template area now displays the name of the template that you chose.

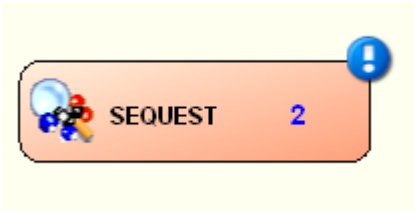
When you open an existing workflow template, some of the nodes in the workspace pane might exhibit a yellow warning symbol, as shown in the example in Figure 64. This symbol indicates that the version of the node used when the template was created has been superseded by a later version in the current Proteome Discoverer application. Delete the node from the workflow, and drag the node with the same name from the Workflow Nodes pane to the workspace pane.

Figure 64. Warning symbol indicating an outdated node version



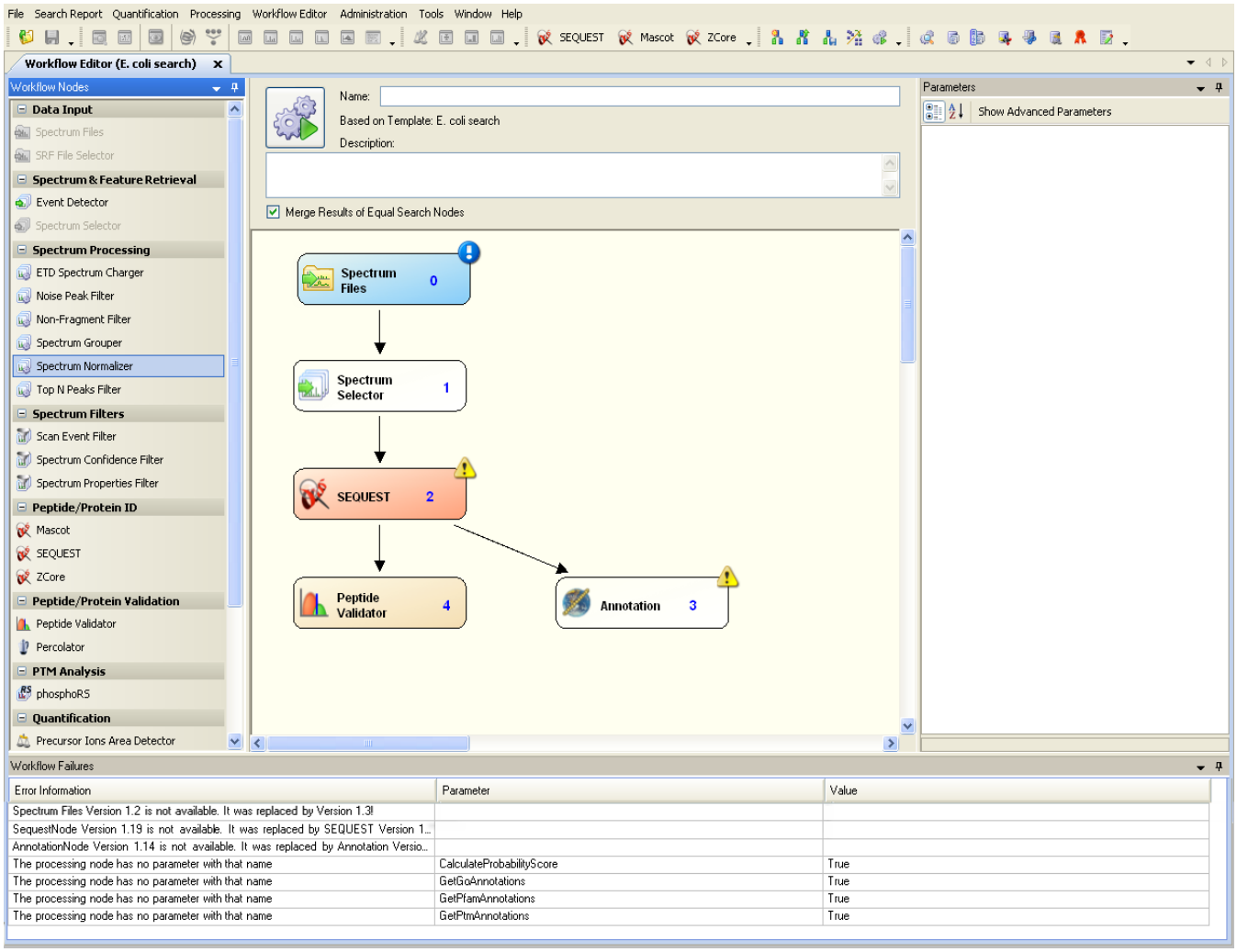
A round blue warning symbol containing an exclamation point, as shown in Figure 65, indicates that one or more of the parameter settings for the node are incorrect or outdated. Click on the node and reset the parameters in the Parameters pane.

Figure 65. Warning symbol indicating incorrect parameter settings



When you use a node that is outdated or has incorrect parameter settings, a Workflow Failures pane opens beneath the Workflow Nodes pane, as shown in Figure 66.

Figure 66. Workflow Failures pane



The Workflow Failures pane contains three columns:

- Error Information: Displays information about the problem that the application encountered in the workflow.
- Parameter: Displays the name of the node parameter that has an erroneous setting.
- Value: Displays the erroneous setting of the node parameter.

When a warning symbol is attached to a node, the Proteome Discoverer application automatically updates the node with the correct version, preserving the previous parameter values in the updated node. It does not include any node parameters that are no longer available and adds any new parameters set to their defaults.

If the Parameter and Value columns indicate a problem with the parameter settings, enter the correct parameter settings in the Parameters pane of the Workflow Editor.

Opening an Existing Workflow from an XML or MSF File

You can open a workflow from an existing MSF or XML file.

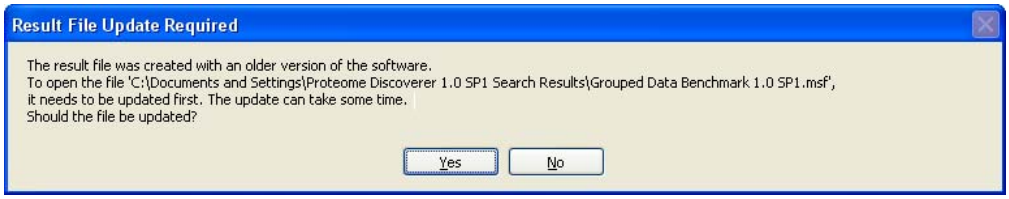
❖ **To open an existing workflow from an XML or MSF file**

1. Choose **Workflow Editor > Import Workflow**.
2. In the Import Workflow dialog box, browse to the XML or MSF file containing the workflow to import, and click **Open**.

   The selected workflow now opens in the Workflow Editor. The Proteome Discoverer application validates parameter settings and uses warning symbols to indicate outdated nodes. It displays error information in the Workflow Failures pane, as shown in Figure 66 on page 111.

   If you selected an MSF file and this file was created with an older version of the Proteome Discoverer application, the message box shown in Figure 67 appears.

   Figure 67. Error message box

3. Click **Yes** to update to the current version.

The Proteome Discoverer application validates parameter settings and displays the selected workflow in the Workflow Editor, using warning symbols to indicate outdated nodes and displaying error information in the Workflow Failures pane.

If the Proteome Discoverer application cannot load the selected MSF file, it displays a message box with information about the issue. It cannot load files that are read-only or invalid, could not be updated, or were created with a newer version of the Proteome Discoverer application.

Deleting an Existing Workflow Template

You can delete an existing workflow template.

❖ **To delete an existing template**

1. Choose **Workflow Editor > Open From Template**.

   The Open Processing Workflow Templates dialog box appears, as shown in the example in Figure 63, listing the available workflow templates.

2. Click the row displaying the name of the template that you want to delete.
3. Click **Remove**.

4. In the Confirm Deletion dialog box, click **Yes**.
5. Click **Remove** again.

Changing the Name and Description of a Workflow Template

You can change the name and the description of a workflow template.

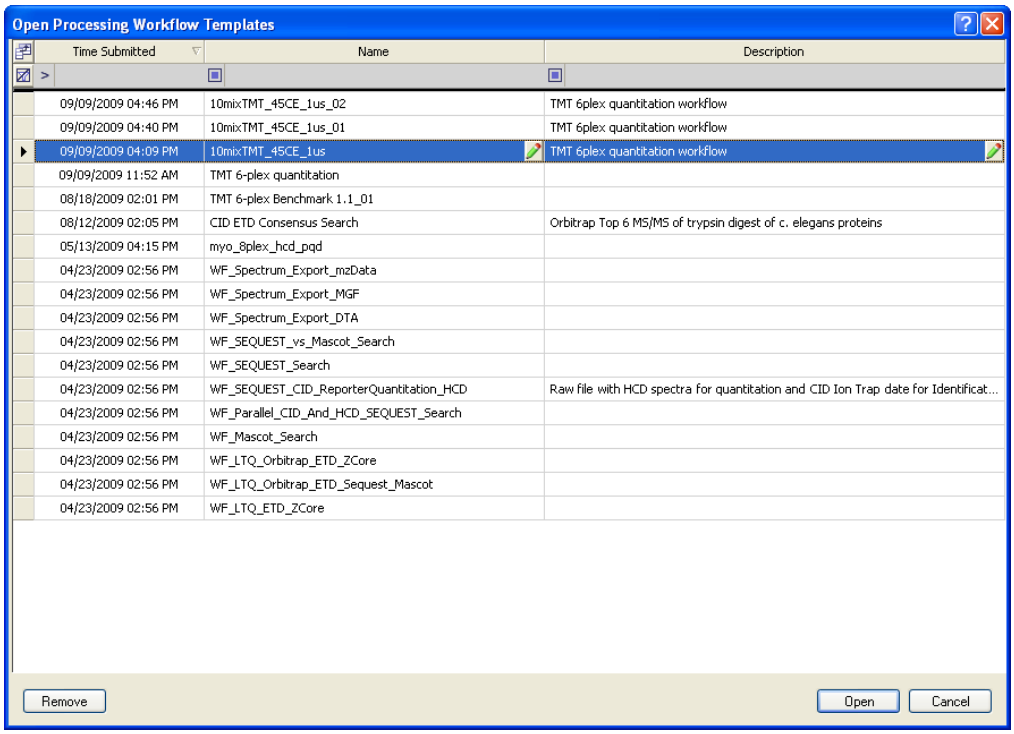
1. Choose **Workflow Editor > Open From Template**.

   The Open Processing Workflow Templates dialog box appears, as shown in the example in Figure 63, listing the available workflow templates.

2. Click the row displaying the name of the template that you want to change.

   A pen icon,  , now appears to the right of the template name and to the right of the template description, as shown in Figure 68.

   Figure 68. Pen icons in the Open Processing Workflow Templates dialog box



3. Click the pen icon,  , and type the new name or the new description.

Importing Raw Data Files in Other Formats into a Workflow

You can import raw data files that were saved in MGF, MZDATA, MZXML, or MZML formats into a workflow.

❖ **To import raw data in MGF, MZDATA, MZXML, or MZML formats**

1. In the Workflow Editor, drag the Spectrum Files node to the workspace pane and select it.
2. In the Parameters pane, click the browse button (**...**) next to the File Name(s) box.
3. In the Select Analysis File(s) dialog box, click **Add Files**.
4. Browse to the location of the MGF, MZDATA, MZXML, or MZML file and select it.
5. Click **Open**.
6. In the Select Analysis File(s) dialog box, click **OK**.
7. Continue with constructing the workflow according to the instructions in "Creating a Search Workflow" on page 91.

Saving a Workflow as an XML Template

To avoid losing any changes, you might want to save a workflow file as an XML template if you intend to transfer it to another computer, another software version, or another person.

❖ **To save a search workflow as an XML template**

1. Choose **Workflow Editor > Export Workflow to XML**.
2. In the Export Workflow Template dialog box, browse to the location where you would like to save the template, type a file name in the File Name box, and click **Save**.

Exporting Spectra

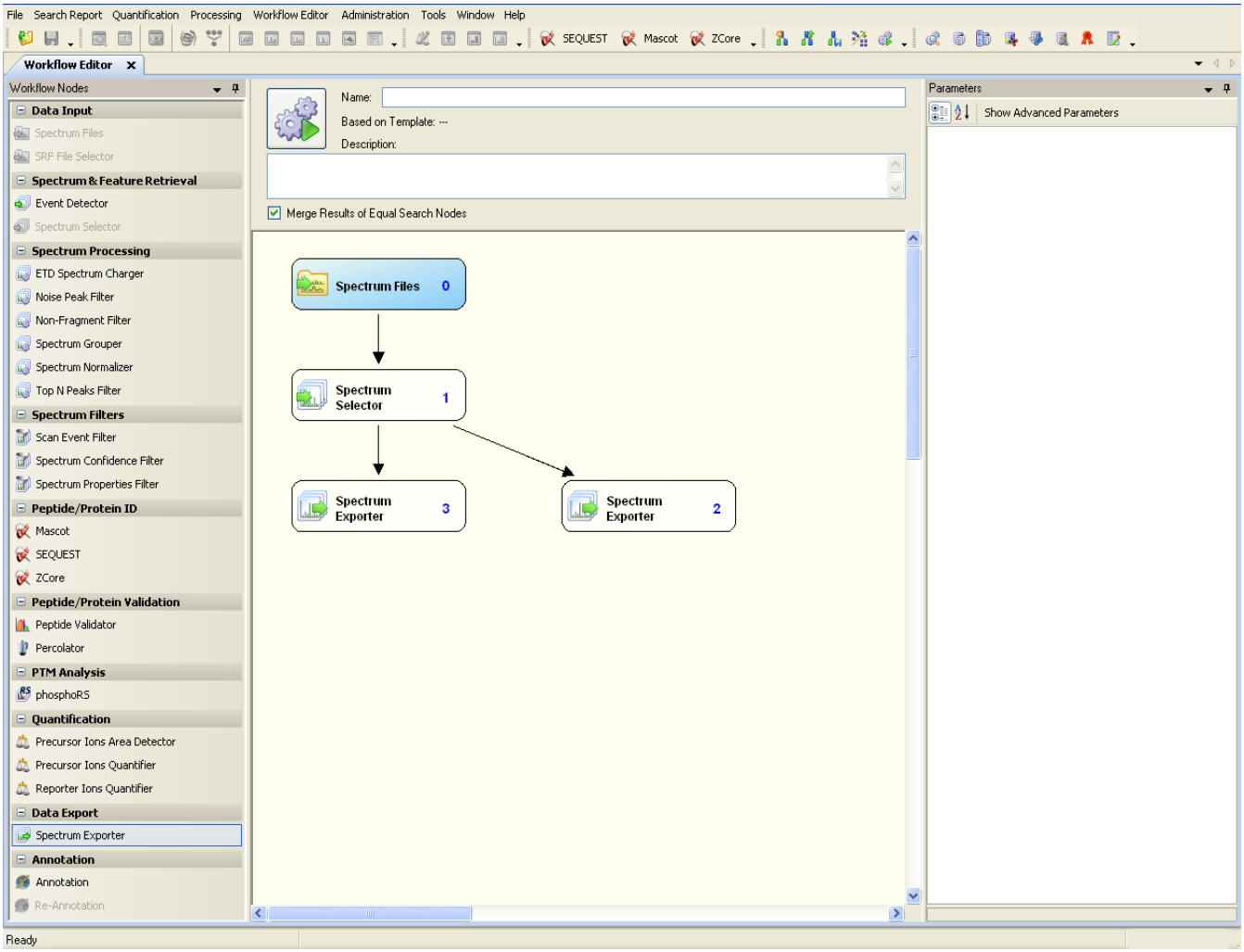
By using the Spectrum Exporter node in your workflow, you can export spectra in the following standard formats:

- Data Archive (DTA): Places the exported spectra into DTA zip files, which are files containing MS^n data for single or grouped scans.
- Mascot Generic Format (MGF): Places the exported spectra into MGF files, which are mass spectral files produced during Mascot analysis. They contain a list of precursor ions, their fragments, and the masses of the fragments.
- MZDATA: Places the exported spectra into MZDATA files, which are common data format files developed by the Human Proteome Organization (HUPO) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line.

- MZXML: These files are standard 2.*x* mass spectrometer data format files developed at the Seattle Proteome Center at the Institute for Systems Biology (ISB) that contain a list of precursor ions, their fragments, and the masses of the fragment.
- MZML: These files are a combination of MZDATA and MZXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the Seattle Proteome Center at the Institute for Systems Biology (ISB). The Proteome Discoverer application supports version 1.1.0.

You can select only one format for each Spectrum Exporter node. To export to multiple formats in a single workflow, you must add more than one Spectrum Exporter node to your workflow, as shown in Figure 69. Set the Export Format parameter on the first Spectrum Exporter node to one format and the Export Format parameter on the next node to another format, and so forth.

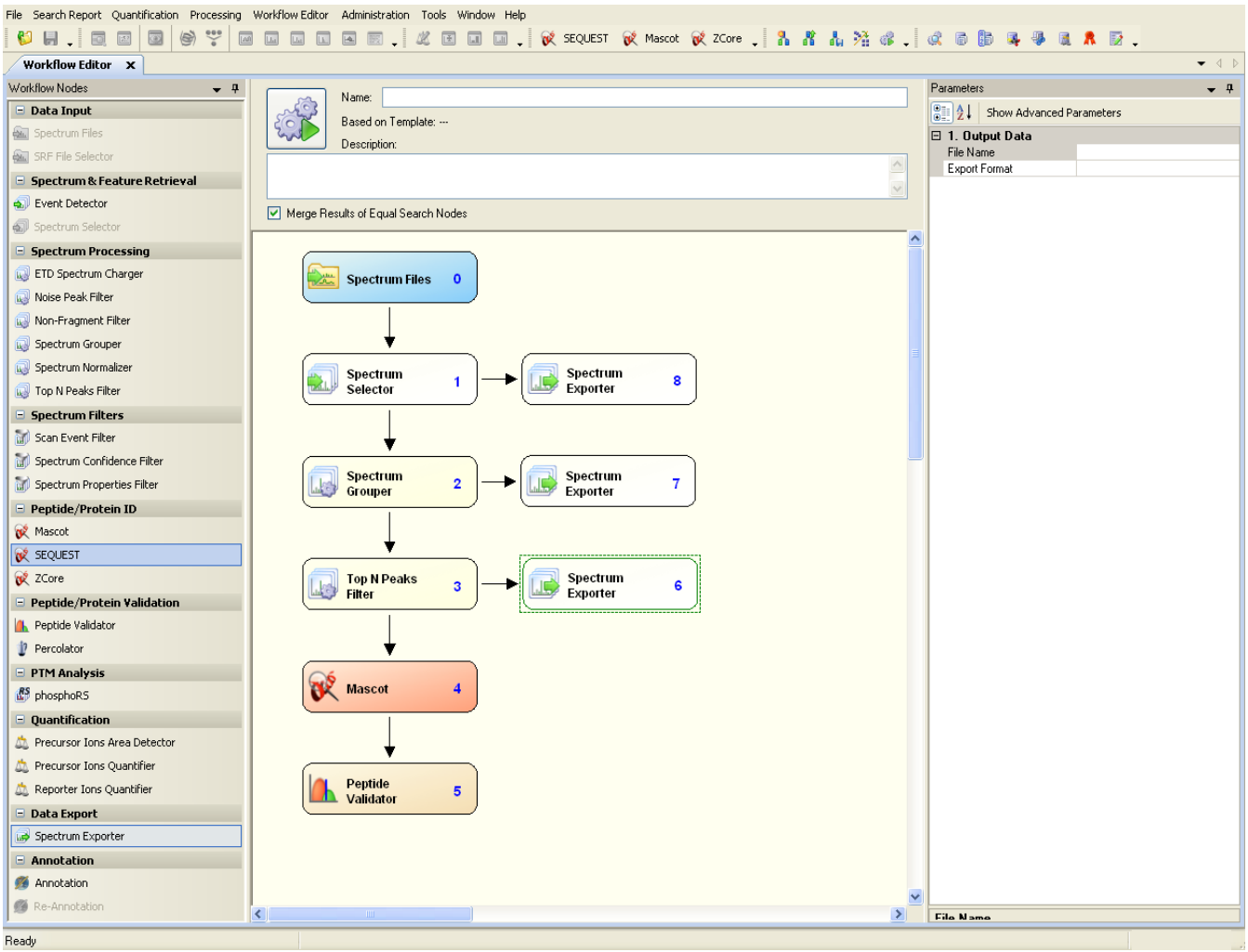
Figure 69. Workflow set to export in two different formats



After starting the export process, the workflow starts like any other workflow processing job. After the application has finished processing the workflow, you can find the output of the Spectrum Exporter node in the same folder as the raw file. The Spectrum Files node specifies the location of the raw file.

You can also attach the Spectrum Exporter node to every node that creates, modifies, or outputs spectra, as shown in Figure 70. For example, you can add the Spectrum Exporter node to the Spectrum Selector node, the Spectrum Filter node, and the Spectrum Processing node. You can use this type of process flow to more closely inspect different spectrum processing steps in a workflow.

Figure 70. Using the Spectrum Exporter node to export spectra from different steps in a workflow



Workflow Nodes

Table 23 briefly describes the nodes available in the Proteome Discoverer Workflow Editor. For more detailed information about these nodes and the parameters that you can set for each node, see "Workflow Editor Nodes" on page 637.

Table 23. Workflow nodes (Sheet 1 of 4)

Workflow node	Function
Data Input	
Spectrum Files	Specifies the name of the raw data file to use for your search. You must use the Spectrum Selector node with the Spectrum Files node. The raw data file can be in either raw, MGF, MZDATA, MZXML, or MZML format. For more information about the Spectrum Files node, see "Spectrum Files Node" on page 638.
SRF File Selector	Specifies the name of the SRF file to use for your search. You can use this node to access a BioWorks SRF file instead of a raw data file for subsequent analysis or format conversion. For more information about the SRF File Selector node, see "SRF File Selector Node" on page 638. For information about importing SRF files from the BioWorks to the Proteome Discoverer application, see "Migrating from BioWorks to the Proteome Discoverer Application" on page 715.
Spectrum And Feature Retrieval	
Event Detector	Clusters isotopes that elute during the same retention time. This node also removes noise and spike signals from the spectra used for further processing. See "Event Detector Node" on page 639.
Spectrum Selector	Specifies the part of the spectrum that you want to search. For more information about the Spectrum Selector node, see "Spectrum Selector Node" on page 640.
Spectrum Processing	
ETD Spectrum Charger	Calculates precursor ion charge states for ETD spectra used for all low-mass-accuracy ETD data. For more information about the ETD Spectrum Charger node, see "ETD Spectrum Charger Node" on page 648.
Noise Peak Filter	Specifies the signal-to-noise thresholds for your search. For more information about the Noise Peak Filter node, see "Noise Peak Filter Node" on page 648.

Table 23. Workflow nodes (Sheet 2 of 4)

Workflow node	Function
Non-Fragment Filter	Removes precursor ion and charge-reduced precursor ion peaks present in Orbitrap or FTMS data spectra. For more information about the Non-Fragment Filter node, see "Non-Fragment Filter Node" on page 649 and "Creating a Search Workflow for Multiple Raw Files from the Same Sample" on page 101.
Spectrum Grouper	Applies a grouping function to the data set. You can use this node to group spectra from MS/MS and MS^3 by setting the Allow MS Order Mismatch option to True. The default is False. To improve the overall fragment ion coverage in alternating experiments (for example, HCD/CID), you can set the Allow Mass Analyzer Mismatch option to True. The default is False. For more information about the Spectrum Grouper node, see "Spectrum Grouper Node" on page 655.
Spectrum Normalizer	Normalizes the intensities of a spectrum so that afterward the largest peak has an arbitrary intensity of 100. This node is useful when you combine two files through the Spectrum Grouper node and one of the original spectra has a very different amplitude than the other. For more information about the Spectrum Normalizer node, see "Spectrum Normalizer Node" on page 658.
Top *N* Peaks Filter	Improves the number of peptides identified with high confidence by reducing the number of peaks in the searched peak lists. For more information about the Top *N* Peaks Filter node, see "Top N Peaks Filter Node" on page 658.
Spectrum Filters	
Scan Event Filter	Divides data from a "mixed-mode" experiment so that each "mode" can be processed independently. For example, if you have a raw file obtained with alternating ETD and CID scans, you might want to analyze the CID portion of the data with SEQUEST and the ETD portion with ZCore. For more information about the Scan Event Filter node, see "Scan Event Filter Node" on page 659.

Table 23. Workflow nodes (Sheet 3 of 4)

Workflow node	Function
Spectrum Confidence Filter	Sets up iterative searches. You can specify score thresholds for the purpose of re-searching all spectra that do not have a peptide identified with at least this score. For more information about the Spectrum Score Filter node, see "Spectrum Confidence Filter Node" on page 664.
Spectrum Properties Filter	Sets thresholds for the data that you pick to analyze. It performs the same function as the Spectrum Selector node, but it is included separately so that you can use separate thresholds in different branches of a branched workflow. For more information about the Spectrum Properties Filter node, see "Spectrum Properties Filter Node" on page 662.
Peptide/Protein ID	
Mascot	Analyzes the data with the Mascot search engine. For more information about the Mascot node, see "Mascot Node" on page 666.
SEQUEST	Analyzes the data with the SEQUEST search engine. For more information about the SEQUEST node, see "SEQUEST Node" on page 671.
ZCore	Analyzes the data with the ZCore search engine. ZCore can only process ETD spectra. For more information about the ZCore node, see "ZCore Node" on page 677.
Peptide/Protein Validation	
Peptide Validator	Calculates the probability that the search algorithm incorrectly included a peptide in a sample. This node is used only for peptides. For more information about the Peptide Validator node, see "Peptide Validator Node" on page 680.
Percolator	Uses semi-supervised machine learning to discriminate correct from incorrect PSMs and calculates accurate statistics such as the q-value (FDR) and posterior error probability to improves the number of confidently identified peptides at a given false discovery rate.

Table 23. Workflow nodes (Sheet 4 of 4)

Workflow node	Function
PTM Analysis	
phospho*RS*	Provides a confidence measure for the localization of phosphorylation in peptide sequences. For more information about the phospho*RS* node, see "phosphoRS Node" on page 691.
Quantification	
Precursor Ions Area Detector	Performs peak area calculation quantification. For more information about the Precursor Ions Area Detector node, see "Precursor Ions Area Detector Node" on page 698.
Precursor Ions Quantifier	Enables isotopically labeled precursor ion quantification. This node is used in precursor labeled experiments, such as SILAC. For more information about the Precursor Ions Quantifier node, see "Precursor Ions Quantifier Node" on page 699.
Reporter Ions Quantifier	Enables isobarically labeled reporter ion quantification. This node is used in iTRAQ and TMT experiments. For more information about the Reporter Ions Quantifier node, see "Reporter Ions Quantifier Node" on page 699.
Data Export	
Spectrum Exporter	Converts the spectrum to DTA, MGF, MZDATA, MZXML, or MZML formats for processing by other means. For more information about the Spectrum Exporter node, see "Spectrum Exporter Node" on page 702.
Annotation	
Annotation	Retrieves GO database information from ProteinCenter and installs it in Proteome Discoverer results files. This information is displayed in the GO IDs column in the MSF file. For more information about the Annotation node, see "Annotation Node" on page 704.
Re-Annotation	Updates existing annotations or annotates existing MSF files that do not yet include annotations. For more information about the Re-Annotation node, see "Re-Annotation Node" on page 704

Using the Proteome Discoverer Daemon Utility

This chapter describes the Proteome Discoverer Daemon utility, which you can use to monitor job execution, perform batch processing, and process Multidimensional Protein Identification Technology (MudPIT) samples. You can select a server to connect to, start workflows, and monitor the execution of jobs on the configured server. Unlike the search wizards, which can only perform searches on one raw data file at a time, the Proteome Discoverer Daemon application can perform multiple searches on multiple raw data files at any given time. It can perform searches on multiple raw data files taken from multiple samples or from one sample. You can run the Proteome Discoverer Daemon application on the command line or in a window interface.

Contents

- Starting the Proteome Discoverer Daemon Application in a Window
- Selecting the Server
- Starting a Workflow
- Creating a Parameter File for Use with the Xcalibur Data System
- Monitoring Job Execution in the Proteome Discoverer Daemon Application
- Logging on to a Remote Server
- Running the Proteome Discoverer Daemon Application from the Xcalibur Data System
- Running the Proteome Discoverer Daemon Application on the Command Line

For information about MudPIT and creating a MudPIT workflow, see "Creating a Search Workflow for Multiple Raw Files from the Same Sample" on page 101.

Starting the Proteome Discoverer Daemon Application in a Window

You can start the Proteome Discoverer Daemon application on the command line or in a window. To run it on the command line, see "Running the Proteome Discoverer Daemon Application on the Command Line" on page 153.

❖ To start the Proteome Discoverer Daemon application in a window

1. Start the Proteome Discoverer Daemon application in Windows by choosing **Start > Programs > Thermo Proteome Discoverer *release_number* > Proteome Discoverer Daemon *release_number*** or by clicking the **Daemon** icon,  , on your desktop.
2. After the Proteome Discoverer Daemon application window appears, connect to a computer that is running the Proteome Discoverer application.

Selecting the Server

The Proteome Discoverer Daemon application can connect to a remote server so that you can perform searches on multiple raw data files from multiple samples or one sample on a remote computer. It can also connect to a local server.

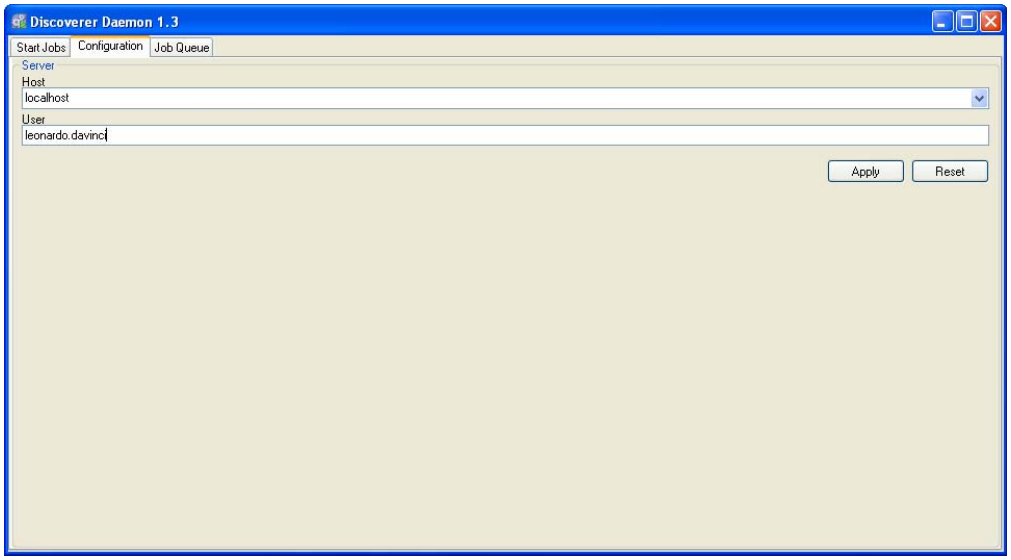
❖ To specify the server to connect to

1. Click the **Configuration** tab in the Proteome Discoverer Daemon application window.
2. From the Host list, select the name of the server that you want to use, or type the server name.

   You must connect the Proteome Discoverer Daemon application to a computer running the Magellan server. Your local host is the default server, that is, the computer that you are working on. To connect to a remote server, see "Logging on to a Remote Server" on page 130.
3. In the User box, type the login name of the server.

   The Configuration page now resembles the illustration in Figure 71.

Figure 71. Configuration page of the Proteome Discoverer Daemon application



4. Click **Apply** to activate the newly entered settings.
5. To return to the previous settings, click **Reset**.

Starting a Workflow

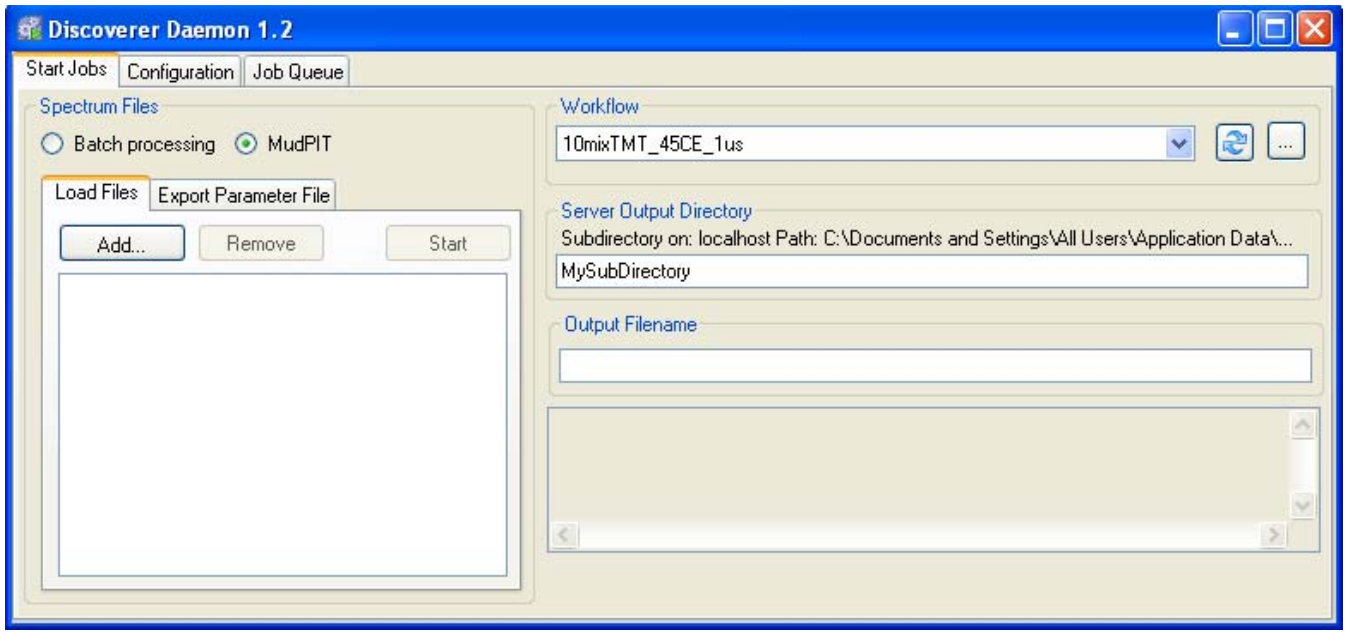
You can start a workflow for batch processing or MudPIT processing.

❖ To start a workflow

1. Click the **Start Jobs** tab.

   The Start Jobs page appears, as shown in Figure 72.

Figure 72. Start Jobs page of the Proteome Discoverer Daemon application



2. Click the **Load Files** tab, if it is not already selected.
3. Click **Add**.
4. In the Open dialog box, locate the file folder containing your raw data, select the spectrum (raw) file or files that you want to load, and click **Open**.

   The selected spectrum file or files appear on the Load Files page.

   To remove a file from the Load Files page, select the file and click **Remove**.

5. To specify the type of processing, select the **Batch Processing** or **MudPIT** option.
   - Batch processing (the default): Executes the workflow once for each spectrum file.
   - MudPIT: Feeds all spectrum files into one workflow.

   When you select the MudPIT option, the Output Filename box becomes available.
6. In the Workflow list, select the workflow template that you want to import.
   - Select the workflow from the Workflow list if it resides on the server that the Proteome Discoverer Daemon application is connected to.

     This workflow must be the one that was saved with the search parameters to be used with the given searches. You cannot modify parameters from the Proteome Discoverer Daemon application itself. Workflow templates that are missing more than the Spectrum File Names parameter do not appear in the Workflow list because the Proteome Discoverer Daemon application cannot complete them.

     –or–
   - Select a valid workflow by clicking the browse button (...) to select the workflow from your local machine.

     If you add workflow templates to the Proteome Discoverer application while the Proteome Discoverer Daemon application is running, click the **Refresh** icon,  , to display the workflow.
7. In the Server Output Directory box, type the name of the directory where you want the original output files placed on the server.

   By default, the Proteome Discoverer Daemon application places this directory under the c:\Documents and Settings\All Users\...\DiscovererDaemon\SpectrumFiles\ directory. If you choose this directory, you must type a file folder name in the Server Output Directory box. You can specify a different directory by choosing Administration > Configuration in the Proteome Discoverer application, clicking Discoverer Daemon in the Server Settings section, and browsing for the location in the New Directory box.
8. If you selected the MudPIT option in the Spectrum Files area, in the Output Filename box, type the name of the output file that you want to store the results of the search in.

   The Start Jobs page should now resemble Figure 73 for batch processing or Figure 74 for MudPIT processing.

Figure 73. Start Jobs page of the Proteome Discoverer Daemon application for batch processing

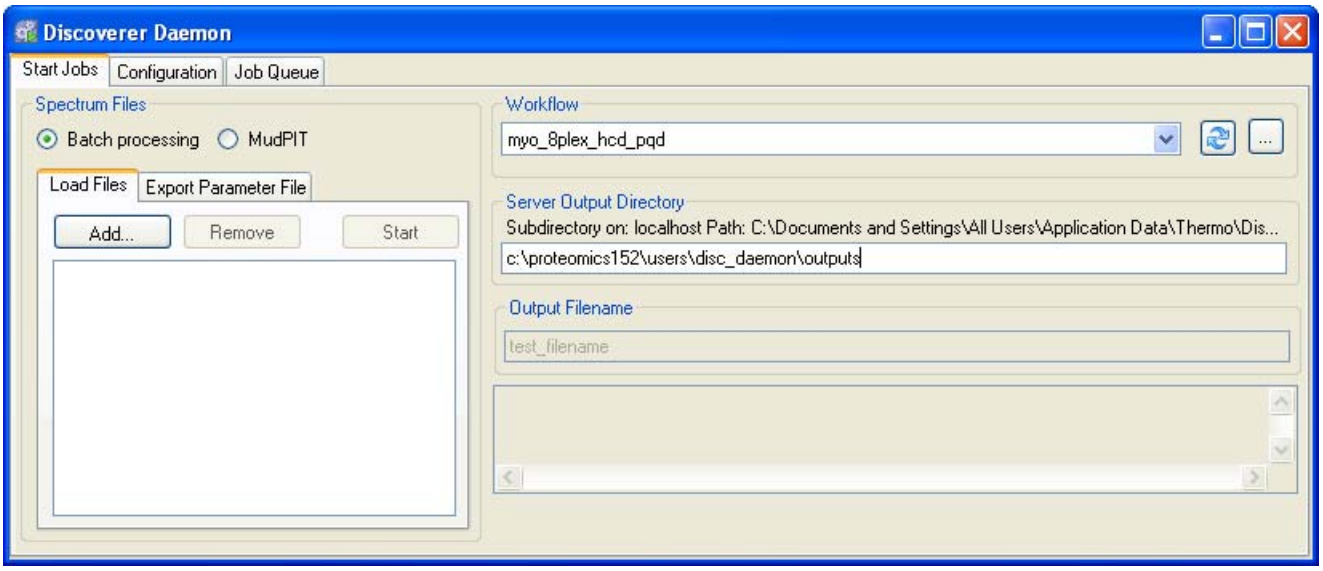
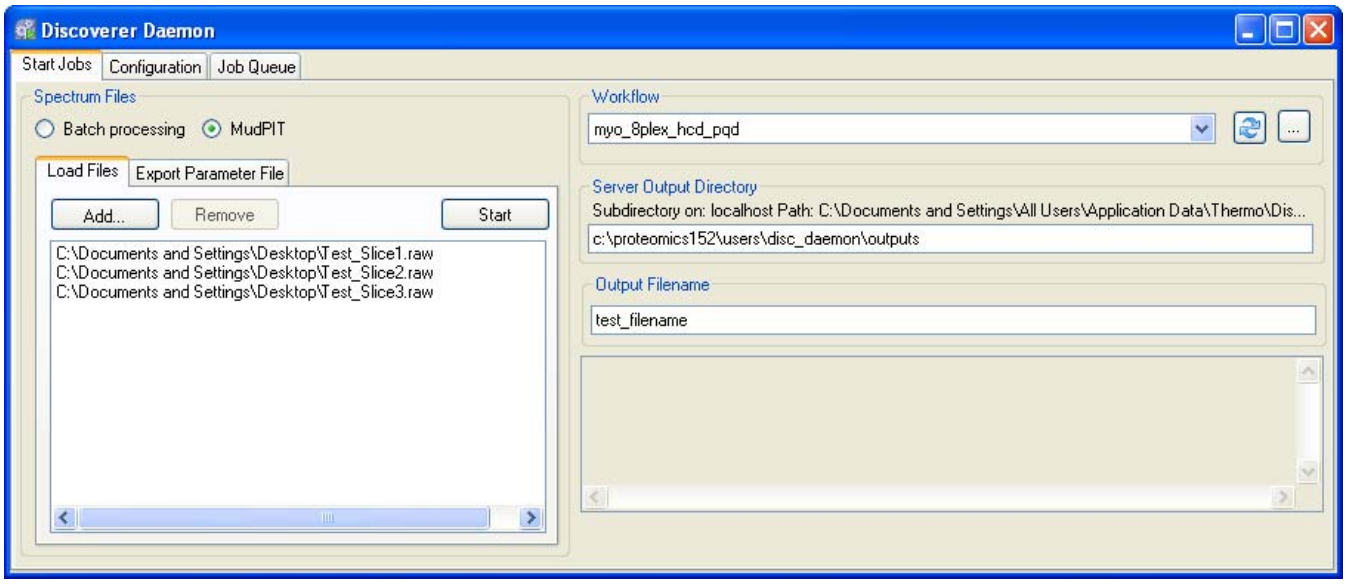


Figure 74. Start Jobs page of the Proteome Discoverer Daemon application for MudPIT processing



9. Click **Start** to execute the job.

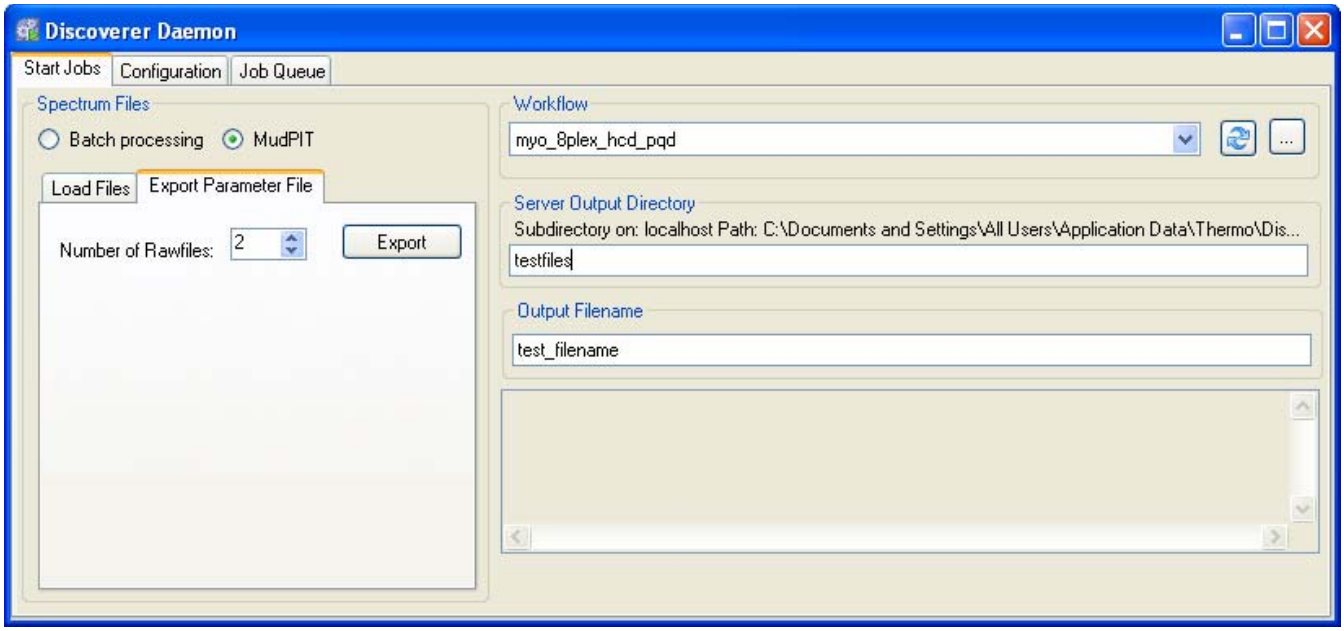
Creating a Parameter File for Use with the Xcalibur Data System

In the Proteome Discoverer Daemon application, you can create a parameter file that you can use to call the Proteome Discoverer Daemon application from the Xcalibur data system. The Proteome Discoverer Daemon application automatically translates the options that you set in the Proteome Discoverer Daemon application interface and in the workflow used for the search into text commands in the parameter file.

❖ To create a parameter file for use with the Xcalibur data system

1. Set up the search according to the instructions in "Starting a Workflow" on page 123. However, you do not have to have files loaded to create a parameter file.
2. Click the **Export Parameter File** tab, shown in Figure 75, on the Start Jobs page.

Figure 75. Export Parameter File page



3. In the Number of Rawfiles box for a MudPIT search, select the number of files that will appear in the Xcalibur Sequence Setup dialog box processing.

   The Number of Rawfiles option is not available when you select batch processing.

4. Click **Export**.

   The Save a Parameter File dialog box appears.

5. Specify the path and name of the parameter file, and click **Save**.

   The Proteome Discoverer application writes the parameter file in .xml format to the directory specified.

To call the Proteome Discoverer Daemon application through the parameter file, see "Running the Proteome Discoverer Daemon Application from the Xcalibur Data System" on page 132.

Monitoring Job Execution in the Proteome Discoverer Daemon Application

You can use the Job Queue page in the Proteome Discoverer Daemon application window to monitor the execution of the jobs that you submit. It performs the same function as the job queue in the Proteome Discoverer interface. For information about the features of the job queue in the Proteome Discoverer interface, see "Job Queue" on page 28.

A progress bar displays the progress of the overall batch processing. This progress bar is only visible if you have started batch jobs.

❖ To monitor the job execution

- Click the **Job Queue** tab of the Proteome Discoverer Daemon application window.

Figure 76 shows the completed job for batch processing, and Figure 77 shows the completed job for MudPIT processing.

Figure 76. Job Queue page of the Proteome Discoverer Daemon application for batch processing

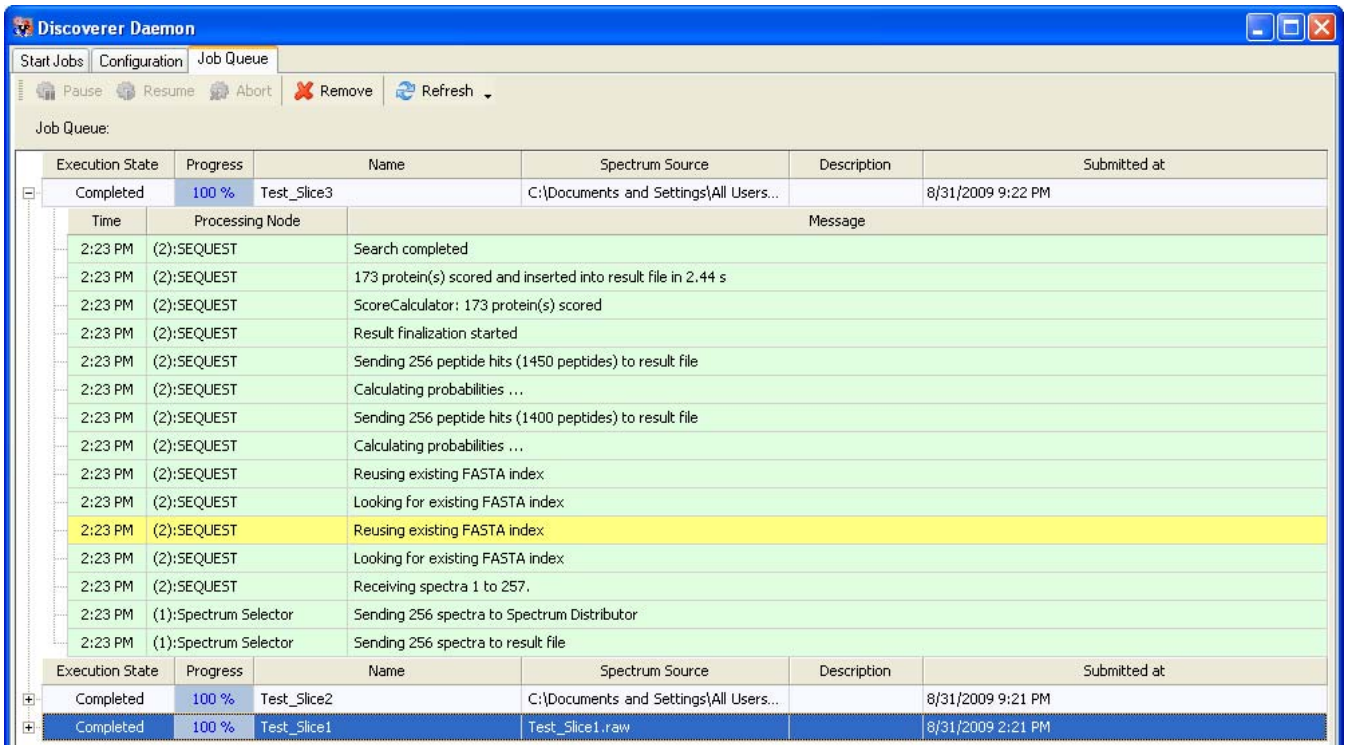
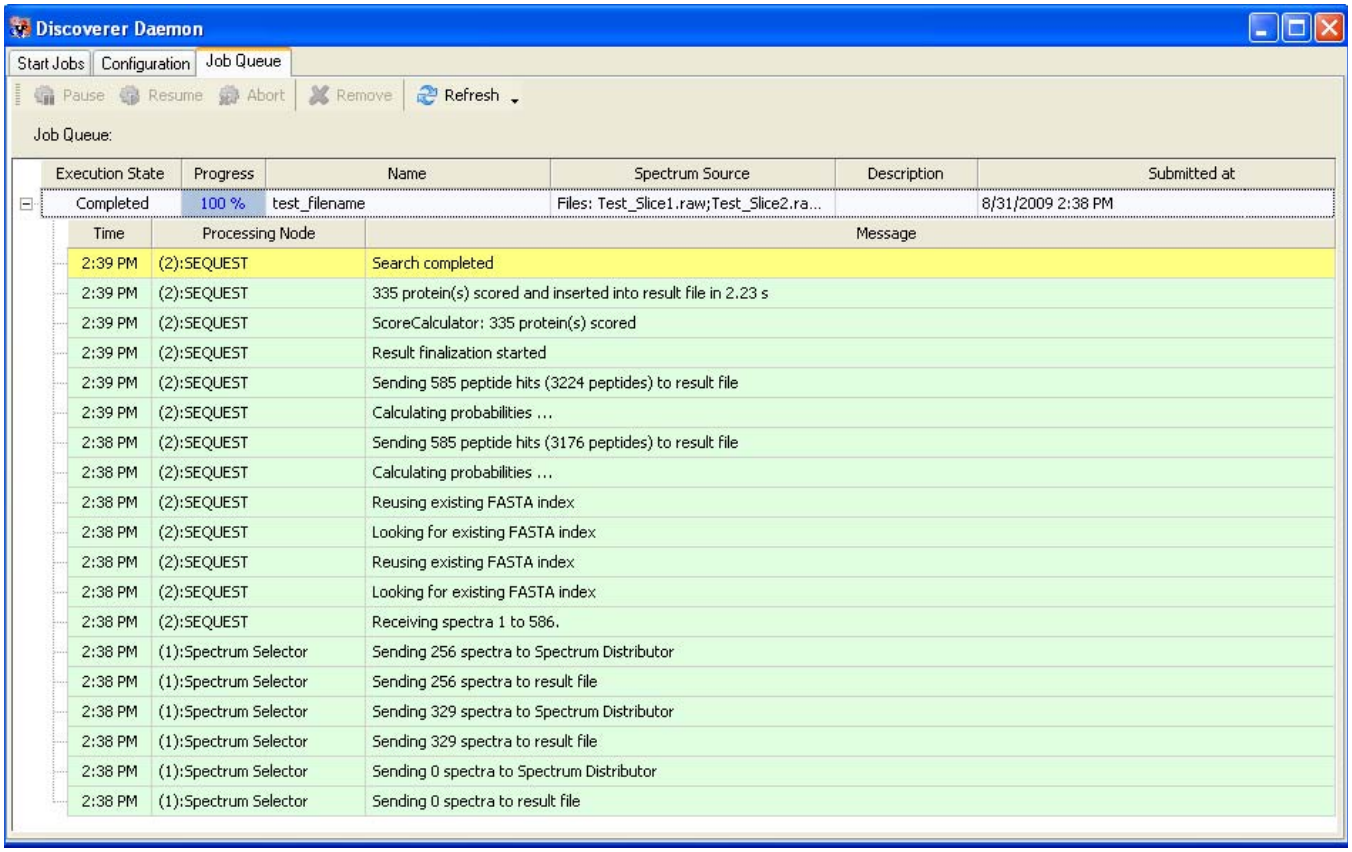


Figure 77. Job Queue page of the Proteome Discoverer Daemon application for MudPIT processing



Proteome Discoverer Daemon Application Window Parameters

The following sections describe the parameters on the pages of the Proteome Discoverer Daemon application window.

Start Jobs Page

Table 24 describes the parameters on the Start Jobs page.

Table 24. Start Jobs page parameters (Sheet 1 of 2)

Parameters	Description
Batch processing	Executes the workflow once for each spectrum file.
MudPIT	Feeds all spectrum files into one workflow.
Load Files	Specifies the files to process and starts the processing.
Add	Opens the Open dialog box so that you can select the spectrum (raw) file or files that you want to load.
Remove	Removes a file or files from the Load Files page.
Start	Executes the job.

Table 24. Start Jobs page parameters (Sheet 2 of 2)

Parameters	Description
Export Parameter File	Creates a parameter file that you can use to call the Proteome Discoverer Daemon application from the Xcalibur data system. The Proteome Discoverer Daemon application automatically translates the options that you set in the application interface and in the workflow used for the search into text commands in the parameter file.
Number of Rawfiles	Specifies the number of raw data files that will appear in the Xcalibur Sequence Setup dialog box for processing. This option is not available for batch processing.
Export	Opens the Save a parameter File dialog box so that you can specify the path and name of the parameter file. When you click Save in this dialog box, the Proteome Discoverer Daemon application writes the parameter file in XML format to the directory that you specified.
Workflow	Specifies the workflow template that you want to import. The template files are listed in alphabetical order on the menu.
Server Output Directory	Specifies the name of the directory where you want the original output files placed on the server.
Output Filename	Specifies the name of the output MSF file in which to store the results of the search from MudPIT samples. This option is available only if you selected the MudPIT option in the Spectrum Files area.

Configuration Page

Table 25 describes the parameters on the Configuration page.

Table 25. Configuration page parameters

Parameters	Description
Host	Specifies the name of the server that you want to use.
User	Specifies the login name of the server.

Job Queue Page

For information about the parameters on the Job Queue page, see Table 3.

Logging on to a Remote Server

The searches started by the Proteome Discoverer application consume memory and can potentially cause the data-acquiring computer to crash and lose the sample in the mass spectrometer. To avoid this outcome, Thermo Fisher Scientific recommends that you connect the Proteome Discoverer Daemon application to a remote computer running the Magellan server before data acquisition.

❖ To log on to a remote server

1. Start the Proteome Discoverer application on the remote machine.
2. If you want to store the output files in a location other than the default, do the following:
   a. Choose **Administration > Configuration > Server Settings > Discoverer Daemon**.

      The PublicFiles folder is the default file displayed in the Current File Directory box, as shown in Figure 78 on page 131.

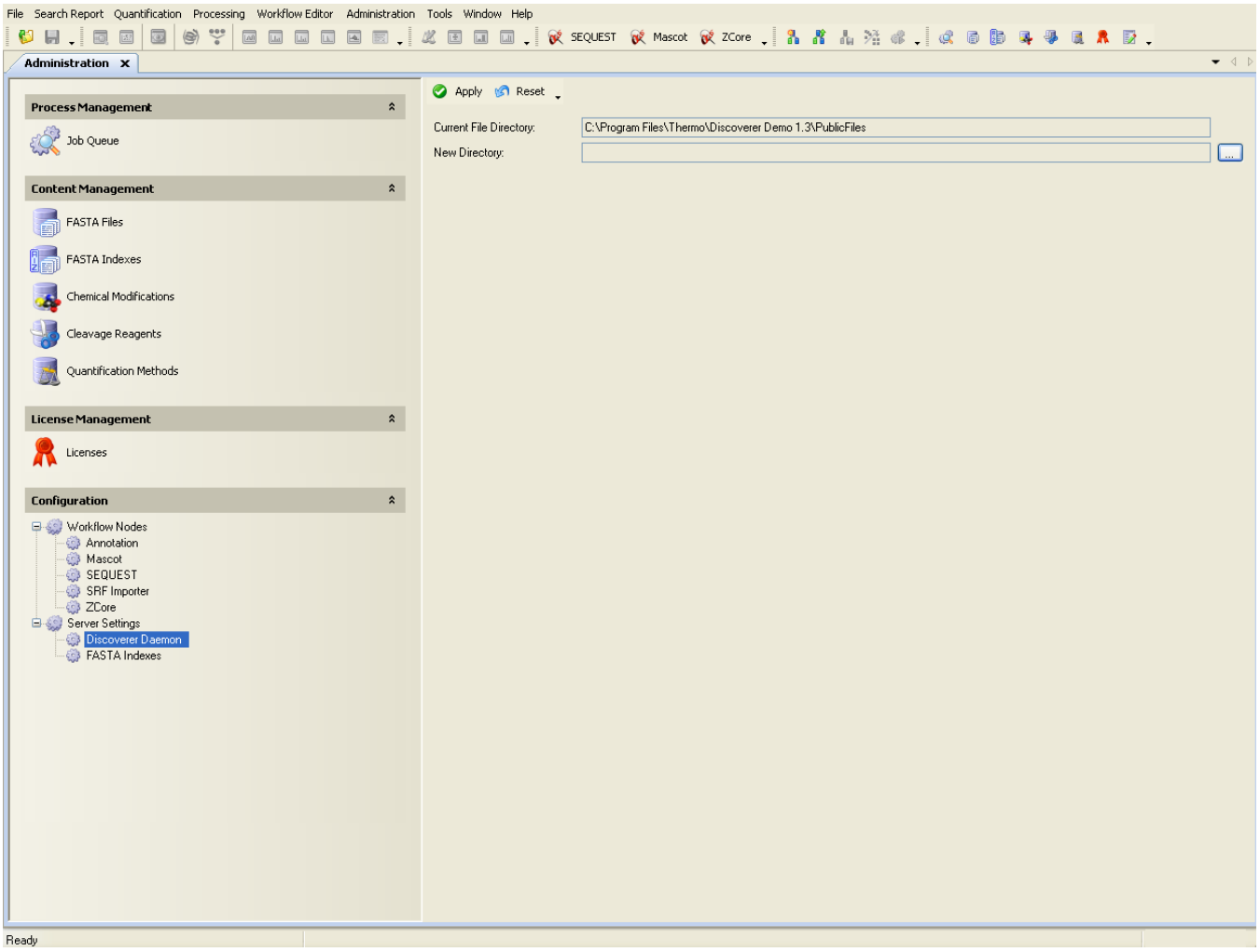
   b. In the New Directory box, browse to the location of the user-named folder in the PublicFiles folder on the server where you want to store the output files.
   c. Click Apply.

   If the directory already exists, it automatically appends the date and an incremental index number to the name.

   If you attempt to create a file other than in the PublicFiles folder in the Current File Directory box, Discoverer Daemon issues a message informing you that the Proteome Discoverer application will apply the change the next time that you start it.

   To return to the default directory, click Reset.

Figure 78. Discoverer Daemon area of the Configuration view



3. Start the Proteome Discoverer Daemon application on the local machine.

   A message box informs you that the Proteome Discoverer Daemon application cannot connect to the server.

4. Click **OK** in the message box.

   The Proteome Discoverer Daemon application opens with the Configuration page selected.

5. In the Host box, type the name of the remote computer.
6. In the User box, type the login name of the remote server.
7. Click **Apply**.

Running the Proteome Discoverer Daemon Application from the Xcalibur Data System

You can use the parameter file created in the Proteome Discoverer Daemon application to call the application from the Xcalibur data system (see "Creating a Parameter File for Use with the Xcalibur Data System" on page 125). For more information about the Xcalibur data system, refer to the *Thermo Xcalibur Acquisition and Processing User Guide,* part number XCALI-97209.

Before You Start

Before you start running the Proteome Discoverer Daemon application from the Xcalibur data system, perform the following steps to ensure that the interface between the Proteome Discoverer Daemon application and the Xcalibur data system is optimal.

❖ To prepare to run the Proteome Discoverer Daemon application from the Xcalibur data system

1. Before you start the Proteome Discoverer Daemon application, install the Proteome Discoverer application on a remote computer to decouple data processing from data acquisition.

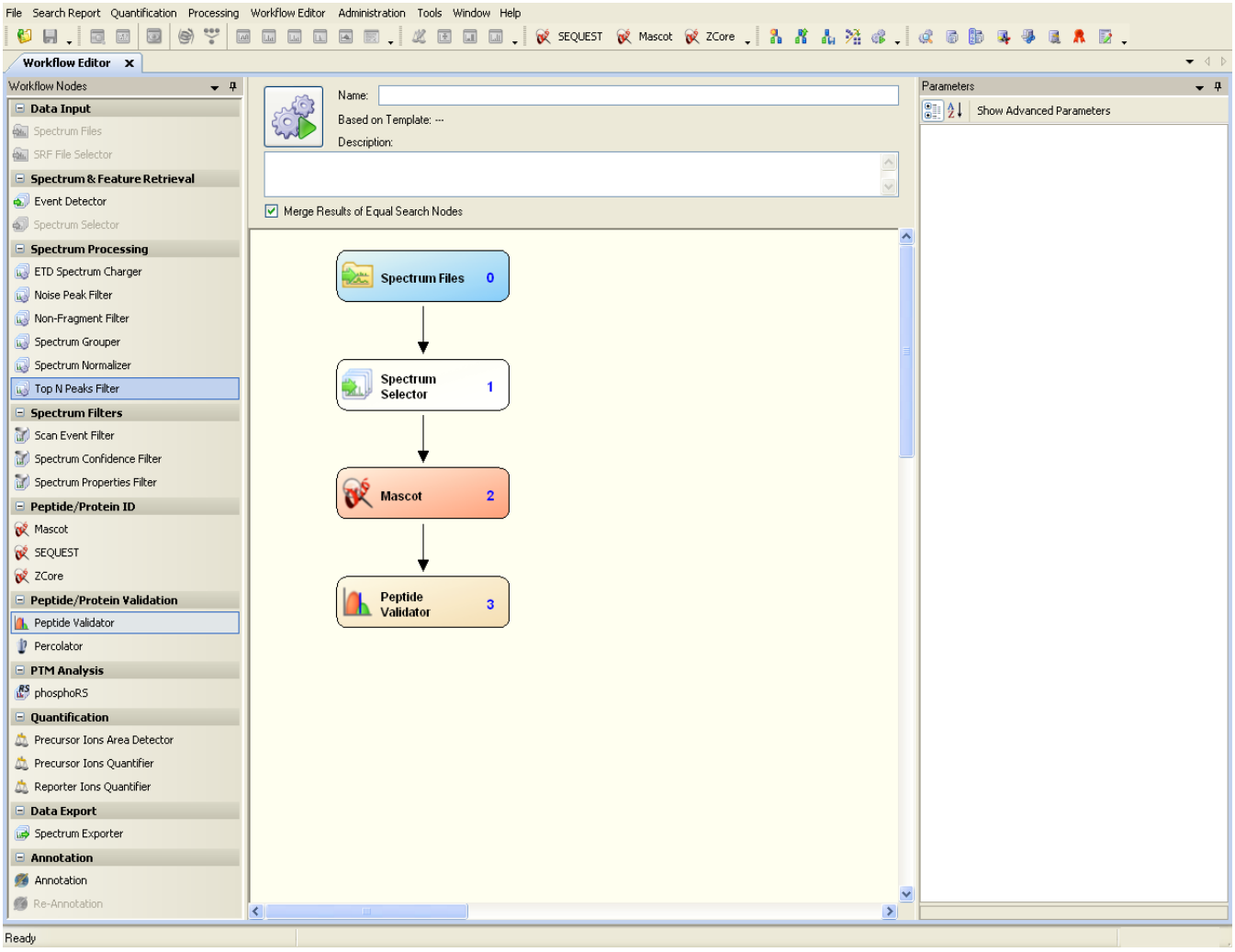
   Thermo Scientific strongly recommends that you perform data analysis and data recording on two different computers to avoid disturbing the data acquisition by resource-consuming data processing.

2. Start the Proteome Discoverer application.

3. Install the Proteome Discoverer Daemon application on the same computer that the Xcalibur data system is running on.

4. In the Proteome Discoverer application, prepare the workflow to be used by the Proteome Discoverer Daemon application, as shown in Figure 79. Save this workflow on the remote server.

Figure 79. Simple workflow used for the samples

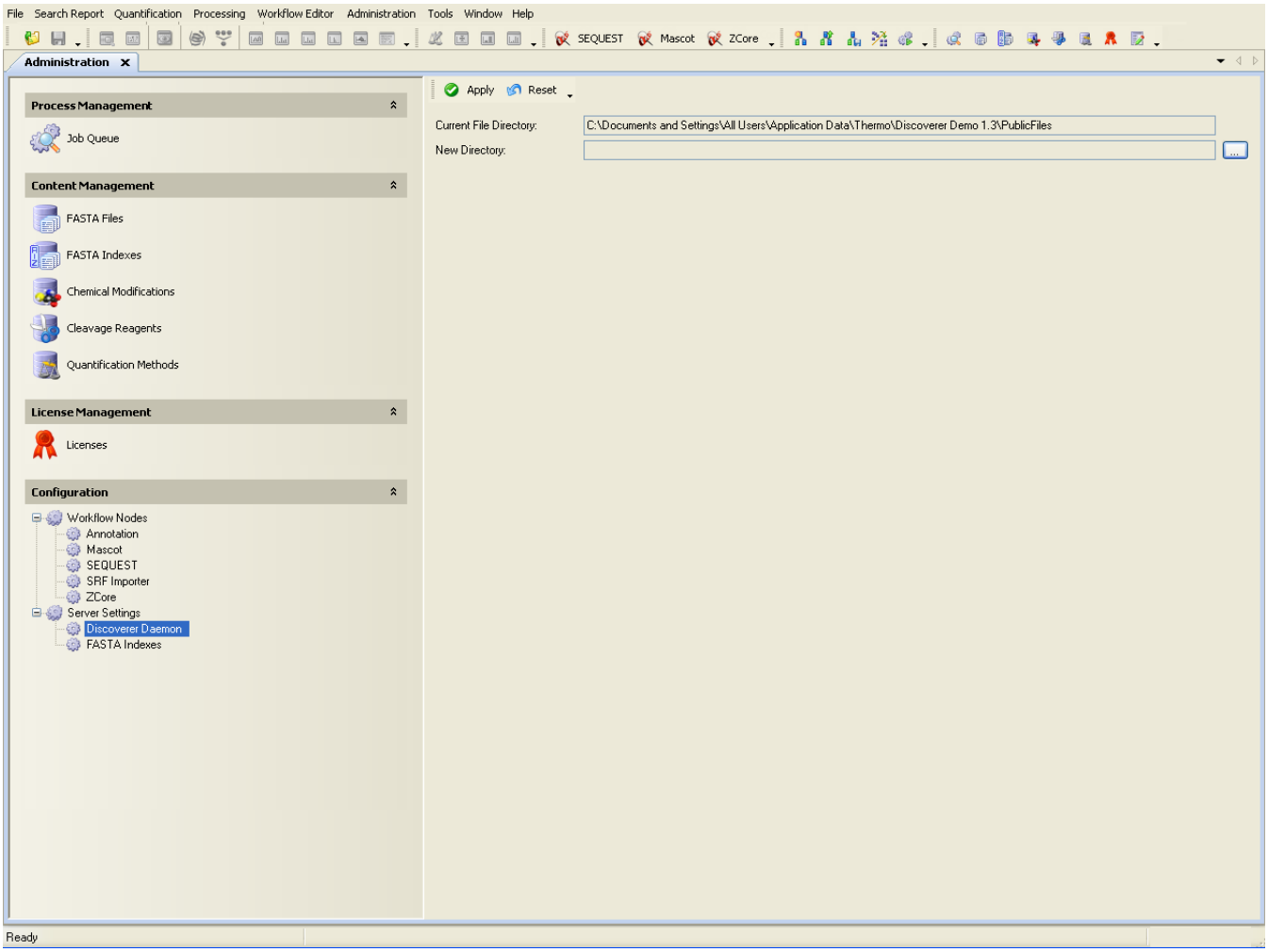


After you install the Proteome Discoverer Daemon application, the Proteome Discoverer application places the directory where it saves the raw files and stores the results in C:\Documents and Settings\All Users\Application data\Thermo\Discoverer\Public Files. This directory might be invisible to you because the C:\Documents and Settings\All Users\Application data directory is hidden. To display hidden directories, choose **Tools > Folder Options > View > Hidden files and folders > Show hidden files and folders** in Windows Explorer.

5. (Optional) To change this directory for easier data access, open the Proteome Discoverer application, choose **Administration > Configuration**, click **Discoverer Daemon** beneath Server Settings in the Configuration area on the left side of the Administration view, and change the directory in the New Directory box, shown in Figure 80.

   The settings are applied after you restart the Proteome Discoverer application.

Figure 80. Changing the directory where results from the Proteome Discoverer Daemon application are stored



Batch Processing Using a Processing Method

To run a batch sample, start the Proteome Discoverer Daemon application on the computer that is running the Xcalibur data system. Connect to the computer that is running the Proteome Discoverer application.

The following procedure assumes that you have a PMD file available.

❖ To generate a processing method

1. Choose **Start > Programs > Thermo Xcalibur > Xcalibur** to start the Xcalibur data system.
2. On the Thermo Xcalibur Home Page, click the **Processing Setup** icon.

3. When the Thermo Xcalibur Processing Setup dialog box appears, do the following:
   a. Create a new processing method or open a previously created one.
   b. Choose **File > Open**, browse to the location of the PMD file, and click **Open**.
   c. Choose **View > View Bar**.

      A toolbar appears on the left side of the dialog box.
   d. Click the **Programs** icon on the left toolbar.
4. When the Programs view of the Thermo Xcalibur Processing Setup dialog box appears, set the samples that the processing method is applied to:
   a. Click the **Enable** box and select the check box that appears.
   b. Accept the defaults in the Std, C, Unk, Other, and Sync boxes or modify them according to your requirements.
   c. Select **Run Program** from the Action list.
   d. Right-click the Program or Macro Name box and choose **Insert Row** from the shortcut menu to open a new row.
   e. Right-click the Program or Macro Name box, choose **Browse** from the shortcut menu, browse to the following executable, and click **Open**.

      C:\Program Files\Thermo\Discoverer\System\Release\DiscovererDaemon.exe

      Note If the following warning appears, click **OK**:

      The file 'DiscovererDaemon' does not exist on this computer.
   f. In the Parameters box, type the location of the parameter file containing the commands that will execute the Proteome Discoverer Daemon application (see "Creating a Parameter File for Use with the Xcalibur Data System" on page 125):

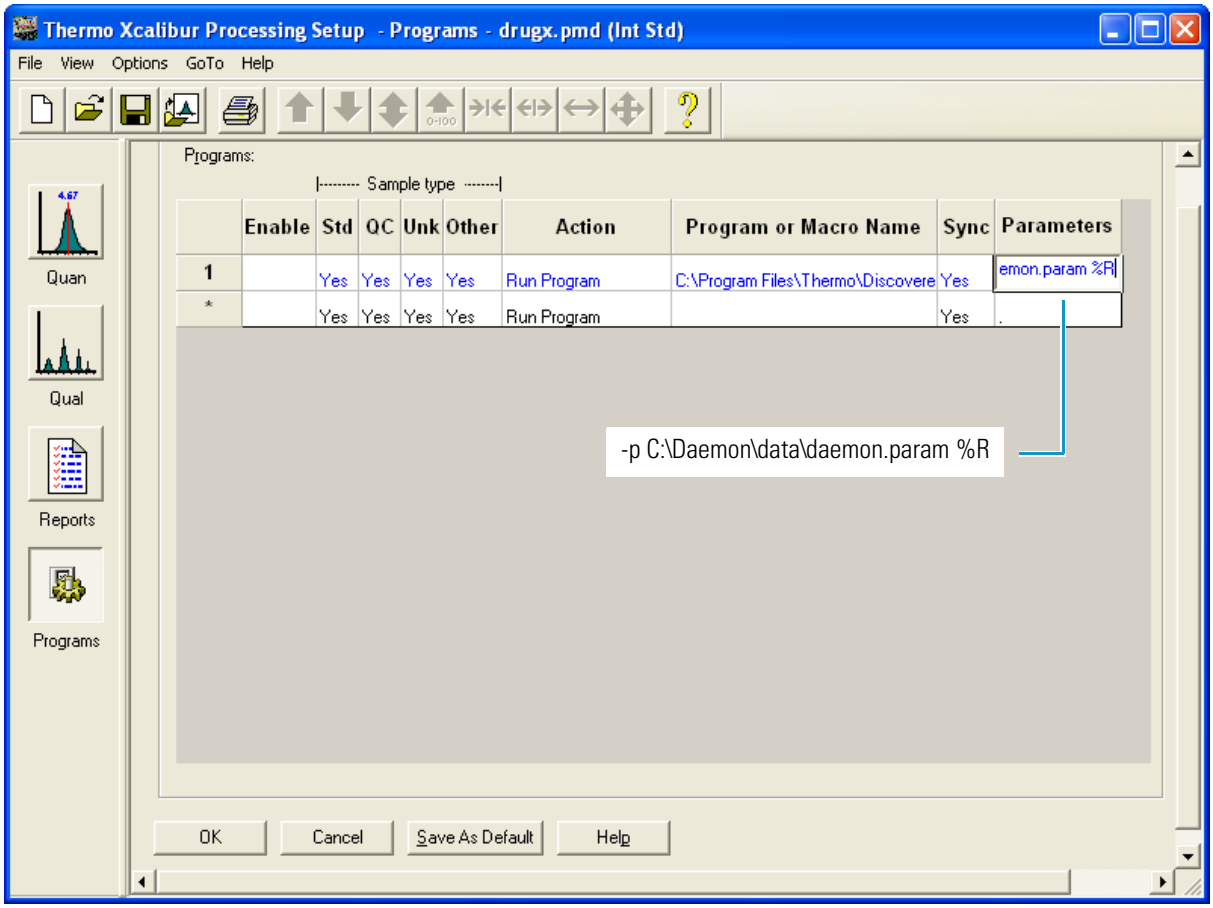
      -p *path_to_parameter_file\parameter_filename* **%R**

      Note If the name of the parameter file contains spaces, you must enclose it in quotation marks, as in this example:

      -p "c:\Xcalibur\methods\batch processing.param" %R

      The Programs view of the Thermo Xcalibur Processing Setup dialog box should now resemble the example in Figure 81.

Figure 81. Completed Thermo Xcalibur Processing Setup - Programs dialog box



g. Activate the row by clicking the Enable box, and then click **OK**.

h. Choose **File > Save**.

The Xcalibur data system now saves the modified PMD file.

i. Close the Processing Setup - Programs dialog box.

Setting Up and Configuring the Mass Spectrometer

To set up and configure the mass spectrometer for data collection, click the **Instrument Setup** icon on the Thermo Xcalibur Home Page and then follow the instructions in the *Thermo Xcalibur Administrator Guide*, part number XCALI-97208.

Setting Up the Experiment Method with the Sequence Editor

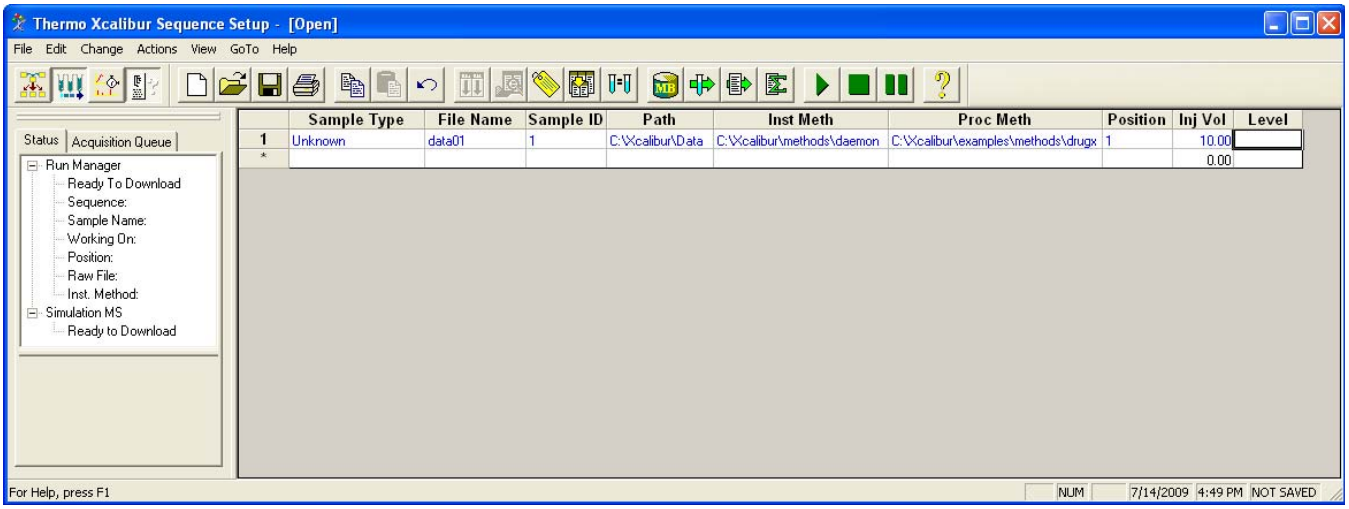
Follow these instructions to set up the experiment method with the Xcalibur Sequence Editor so that the Xcalibur data system can collect the raw data from the mass spectrometer.

❖ To set up the experiment method and run the Proteome Discoverer Daemon application

1. On the Thermo Xcalibur Home Page, click the **Sequence Setup** icon.
2. When the Thermo Xcalibur Sequence Setup dialog box appears, set the parameters for the collection of raw data from the mass spectrometer.
3. Click the **Sample Type** box and select the type of sample from the list.
4. Right-click the File Name box, and enter the name of the file in which you want to store the raw data.
5. In the Sample ID box, use the default or modify it according to your requirements.
6. Right-click the Path box, choose **Browse** from the shortcut menu, select the path of the file where you want to store the raw data, and click **OK**.
7. Right-click the Inst Meth box, choose **Browse** from the shortcut menu, select the instrument method used to collect the data in the raw file, and click **Open**.
8. If the Proc Meth column does not appear, click the **Column Arrangement** icon to add one to the grid.
9. Right-click the Proc Meth box, choose **Browse** from the shortcut menu, select the PMD file that specifies the processing method, and click **Open**.
10. In the Position box, use the default or modify it according to your requirements.
11. In the Inj Vol box, enter the injection volume. For more information about injection volume, refer to the *Thermo Xcalibur Acquisition and Processing User Guide,* part number XCALI-97209.

    The completed Thermo Xcalibur Sequence Setup dialog box now resembles Figure 82.

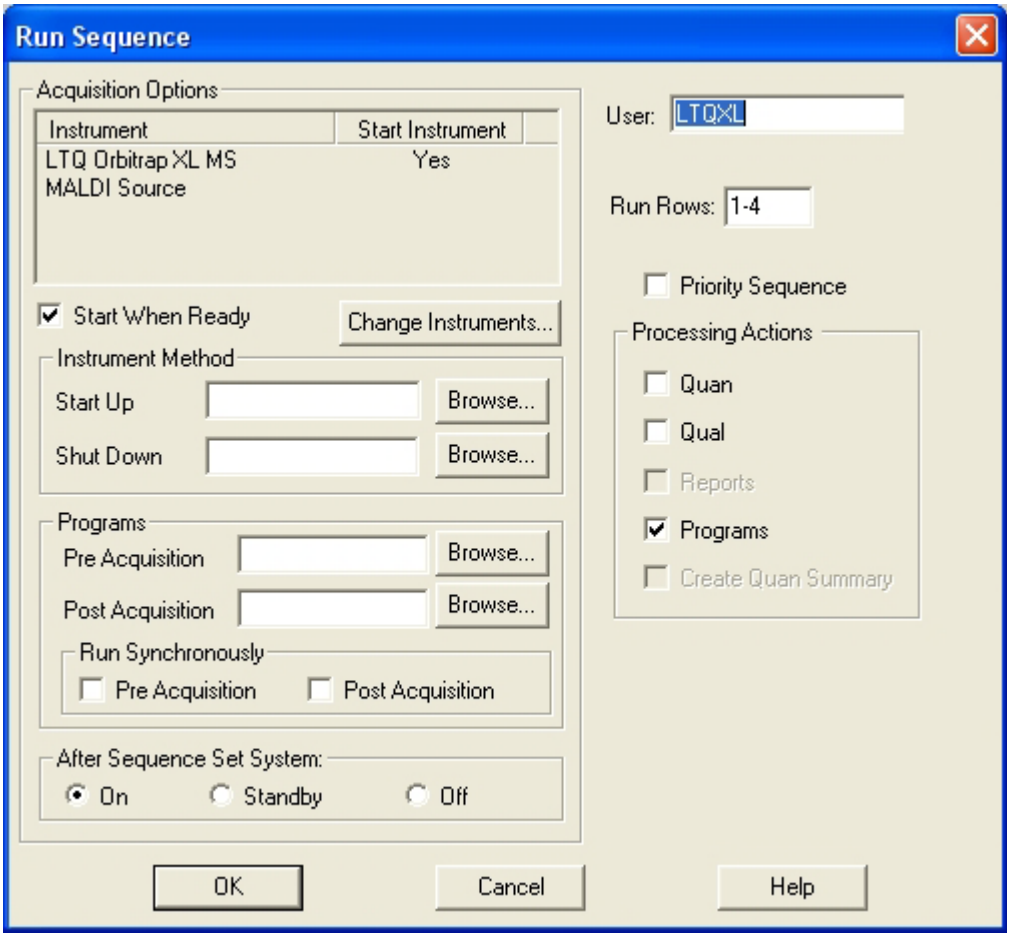
Figure 82. Thermo Xcalibur Sequence Setup dialog box



12. Choose **Actions > Run Sequence** or click the **Run Sequence** icon,  .

    The Run Sequence dialog box appears, as shown in Figure 83.

    Figure 83. Run Sequence dialog box



13. Click **OK**.

The Xcalibur data system now collects the raw data from the mass spectrometer. During the measurement, it does no post-processing. After it finishes acquiring the data, it calls the Proteome Discoverer Daemon application.

After the Xcalibur data system measures all the samples, it calls the processing method for each sample and sends the raw data files to the Proteome Discoverer application, which stores them and the MSF file in the server output directory specified in the Server Output Directory box of the Export Parameter File page of the Start Jobs page. All the search results of the batch processing are stored in the same folder. If the same directory name is used for the results of another batch process, the date and an incrementing index number is appended to the folder name.

Selecting a Single Sample

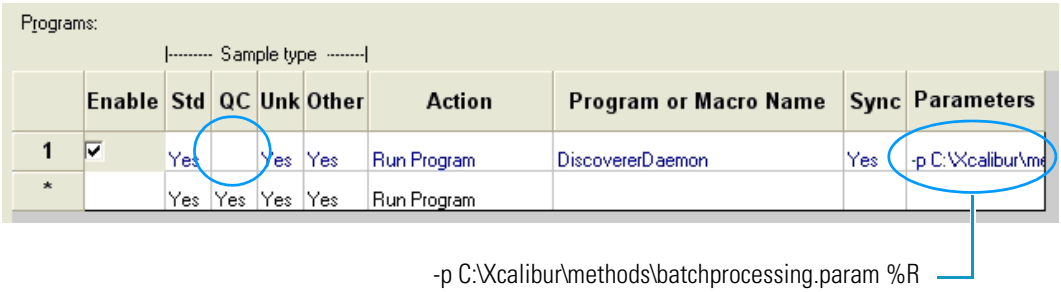
You can apply the processing method to a predefined sample set. You can selectively apply the method to four samples types: Std, C, Unk, and Other.

- To prepare the processing method that excludes all C samples from post-processing
- To use different processing methods in one sequence

❖ To prepare the processing method that excludes all C samples from post-processing

1. Open the processing method explained in "Batch Processing Using a Processing Method" on page 134 by right-clicking the Proc Meth cell in the sequence editor.
2. Modify the processing method by clearing C in the sample type section, as shown in Figure 84.

Figure 84. Modifying the processing method by clearing C samples



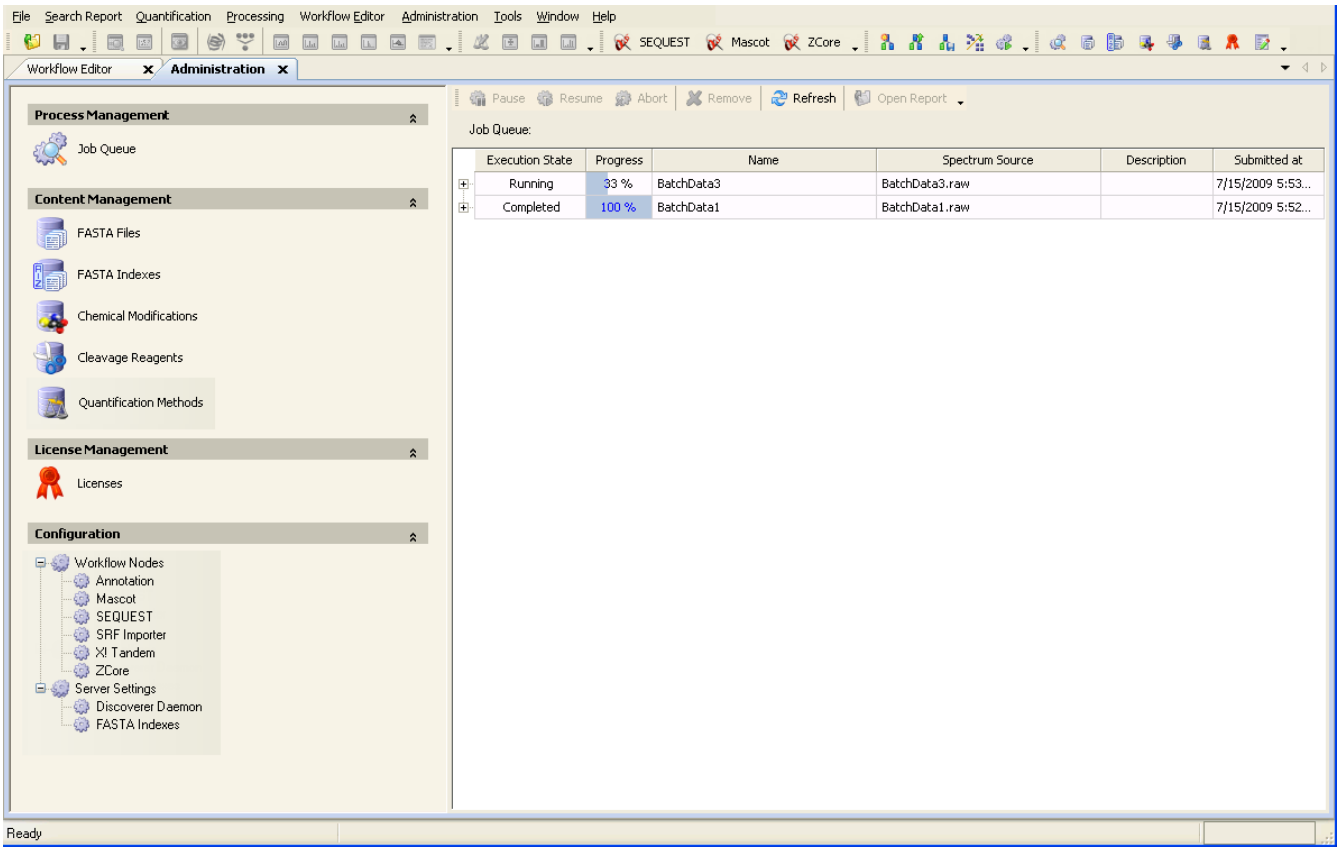
3. Save the method and modify the sequence in the sequence editor by changing the sample type in line 2 to **C**, as shown in Figure 85.

Figure 85. Marking sample two as C in the modified sequence

	File Name	Path	Inst Meth	Position	Proc Meth	Sample Type
1	BatchData1	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	J7	C:\Xcalibur\methods\PDD_batchprocessing	Unknown
2	BatchData2	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	J8	C:\Xcalibur\methods\PDD_batchprocessing	QC
3	BatchData3	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	J9	C:\Xcalibur\methods\PDD_batchprocessing	Unknown
*						

After the samples are recorded, the Xcalibur data system sends the raw data files to the Proteome Discoverer application. Figure 86 shows the output of the Proteome Discoverer job queue associated with this sample. In this example, the Proteome Discoverer application processes only the BatchData1 and BatchData3 raw files.

Figure 86. Processing the samples in the Proteome Discoverer application

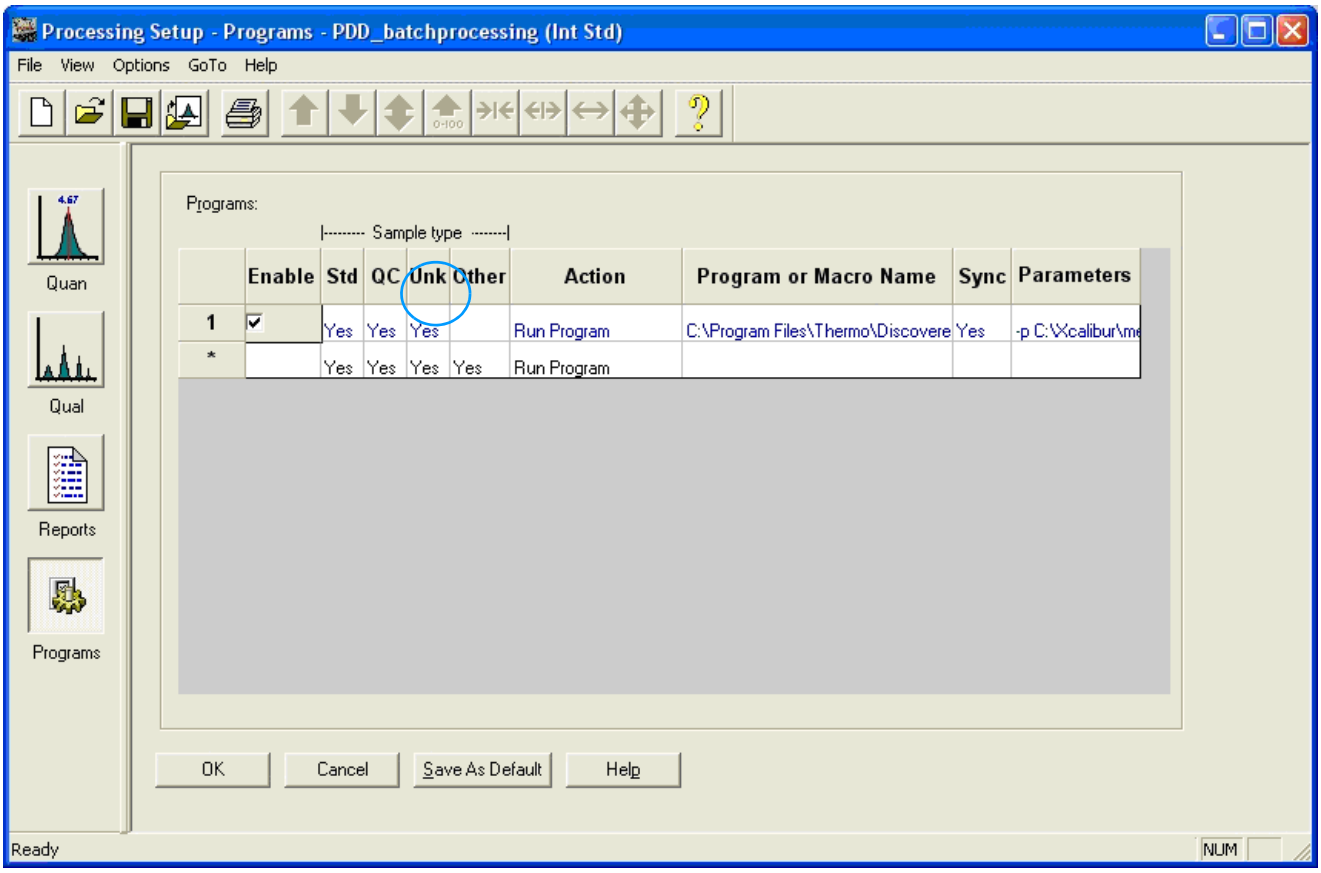


When you clear the Other sample type in the processing setup, the Xcalibur data system skip samples marked as blank in the sequence editor, as shown in Figure 87 and Figure 88.

Figure 87. Sequence with two blank samples

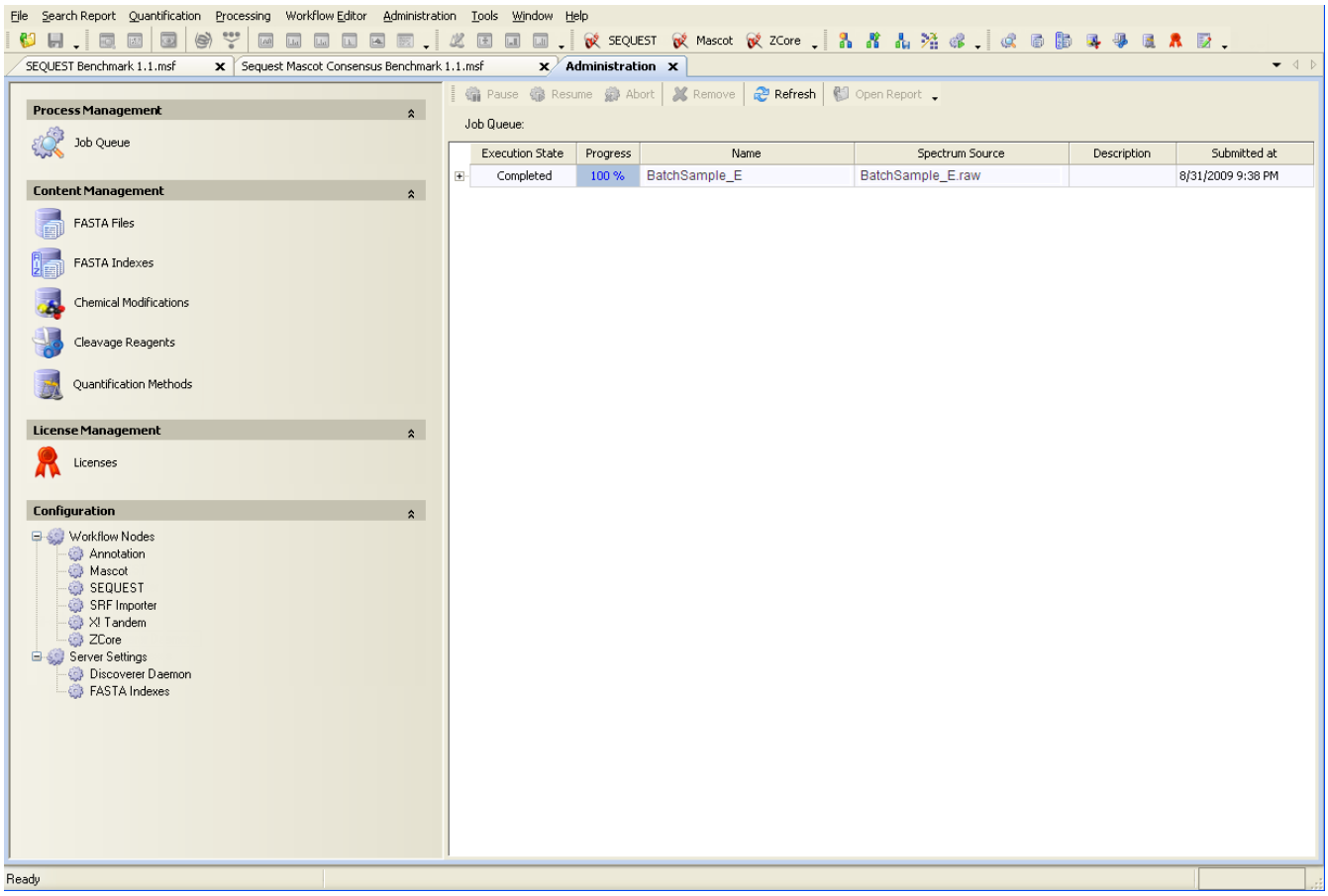
	File Name	Path	Inst Meth	Position	Proc Meth	Sample Type
1	BatchSample_D	C:\Xcalibur\Data	C:\Xcalibur\methods\top50plu	H9	C:\Xcalibur\methods\PDD_batchprocessing	Blank
2	BatchSample_E	C:\Xcalibur\Data	C:\Xcalibur\methods\top50plu	H10	C:\Xcalibur\methods\PDD_batchprocessing	QC
3	BatchSample_F	C:\Xcalibur\Data	C:\Xcalibur\methods\top50plu	H11	C:\Xcalibur\methods\PDD_batchprocessing	Blank
*						

Figure 88. Skipping blank samples in modified processing method



With the sequence shown in Figure 88, the Proteome Discoverer application processes only the BatchSample_E (the sample marked as C), as shown in Figure 89.

Figure 89. Processing only the C samples



❖ To use different processing methods in one sequence

To use different processing methods in one sequence, you must set up the sequence with the None bracket type in the sequence template.

You need two different post-processing methods. For example, the sequest.pmd method runs the Proteome Discoverer Daemon application with a parameter file containing a simple SEQUEST workflow, and the export.pmd method runs the Proteome Discoverer Daemon application with an export workflow.

1. Choose **File > New** in the sequence setup.

   The New Sequence Template dialog box opens.

2. Enter the appropriate values in each of the boxes.

3. In the Bracket Type area, select the **None** option, as shown in Figure 90.

   With this bracket type, you can change the processing methods individually for each sample.

Figure 90. Using the bracket type to change processing methods

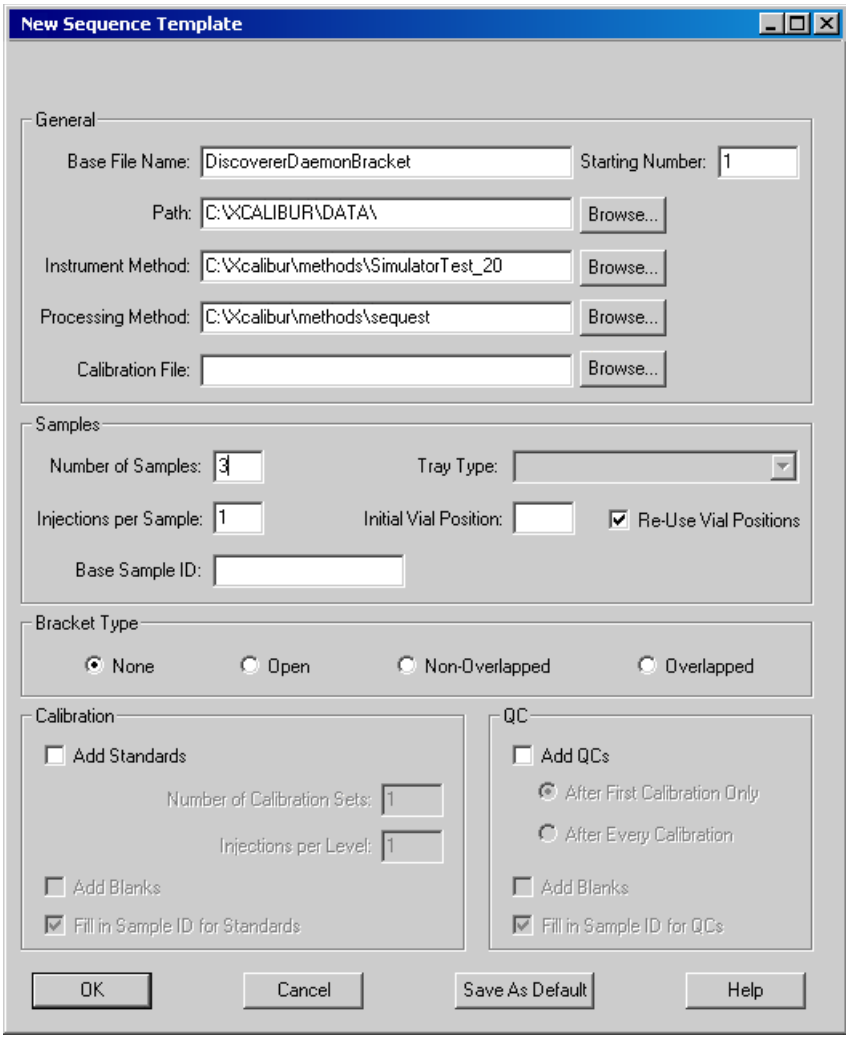


Figure 91 shows a sequence using two different post-processing methods.

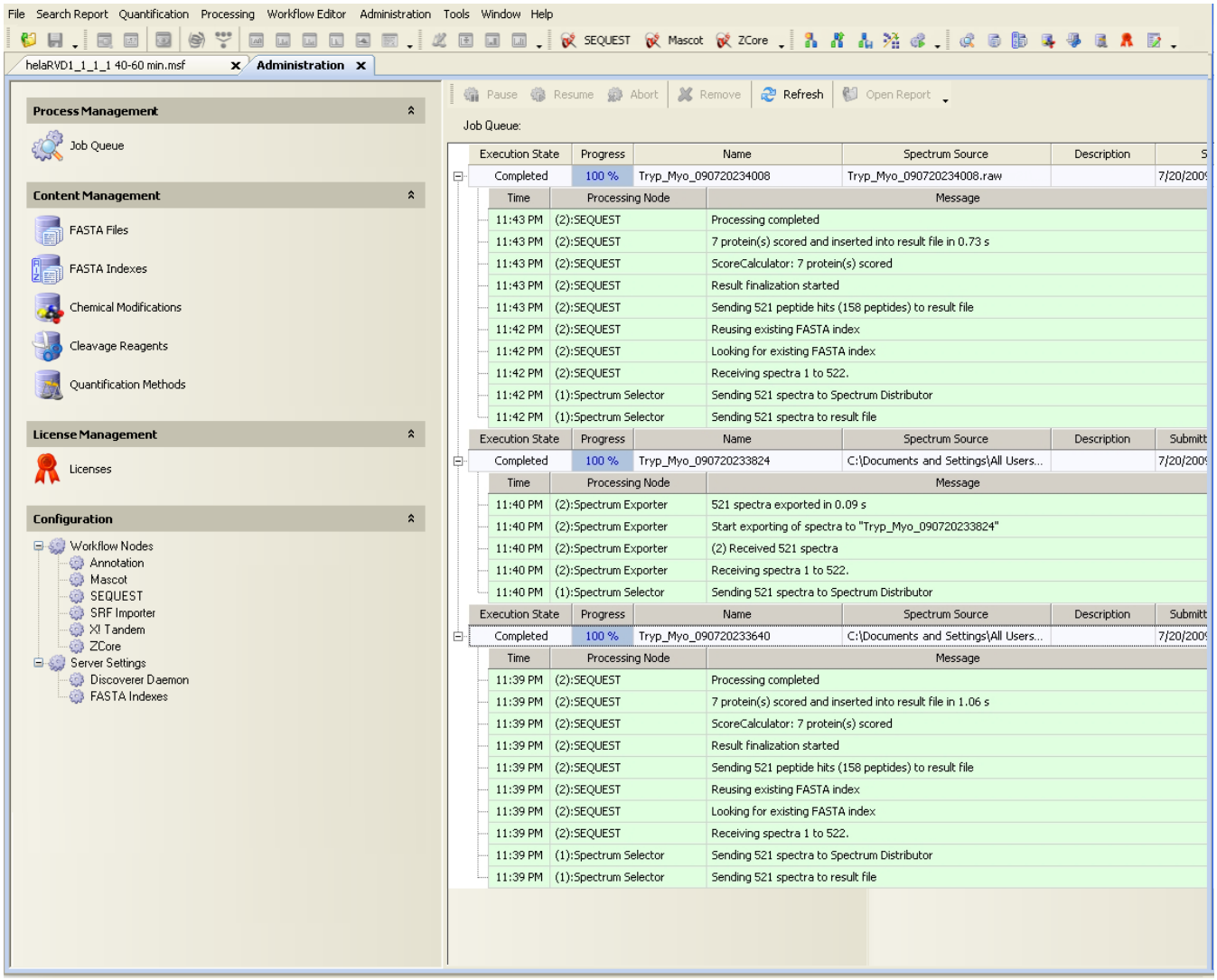
Figure 91. Sequence with two different processing methods

	Sample Type	File Name	Path	Inst Meth	Proc Meth	Position	Inj Vol	Level
1	Unknown	Tryp_Myo	C:\XCALIBUR\[	C:\Xcalibur\mel	C:\Xcalibur\methods\sequest	1	10.00	
2	Unknown	Tryp_Myo	C:\XCALIBUR\[	C:\Xcalibur\mel	C:\Xcalibur\methods\export	2	10.00	
3	Unknown	Tryp_Myo	C:\XCALIBUR\[	C:\Xcalibur\mel	C:\Xcalibur\methods\sequest	3	10.00	
*							0.00	

4. Click **OK**.

In this example, the Xcalibur data system starts two different workflows (performing a SEQUEST search and exporting a raw file) for the recorded raw data files in the Proteome Discoverer application, as shown in Figure 92.

Figure 92. Two workflows in job queue started by two different processing methods



Batch Processing by Using a Post-Acquisition Method

You can perform batch processing by using different processing methods for different samples. However, the processing method does not run synchronously with data recording, and editing the processing method is complicated. For quick synchronous processing of the same workflow, you can use the Proteome Discoverer Daemon application as a post-acquisition method in the Run Sequence dialog box.

Note Using the post-acquisition method with the Proteome Discoverer Daemon application does not work with the Xcalibur 2.1.0 data system.

To use the Proteome Discoverer Daemon application in the Run Sequence dialog box, you do not need a processing method. Figure 93 shows the sequence setup without a processing method.

Figure 93. Sequence used to start batch processing in the Run Sequence dialog box

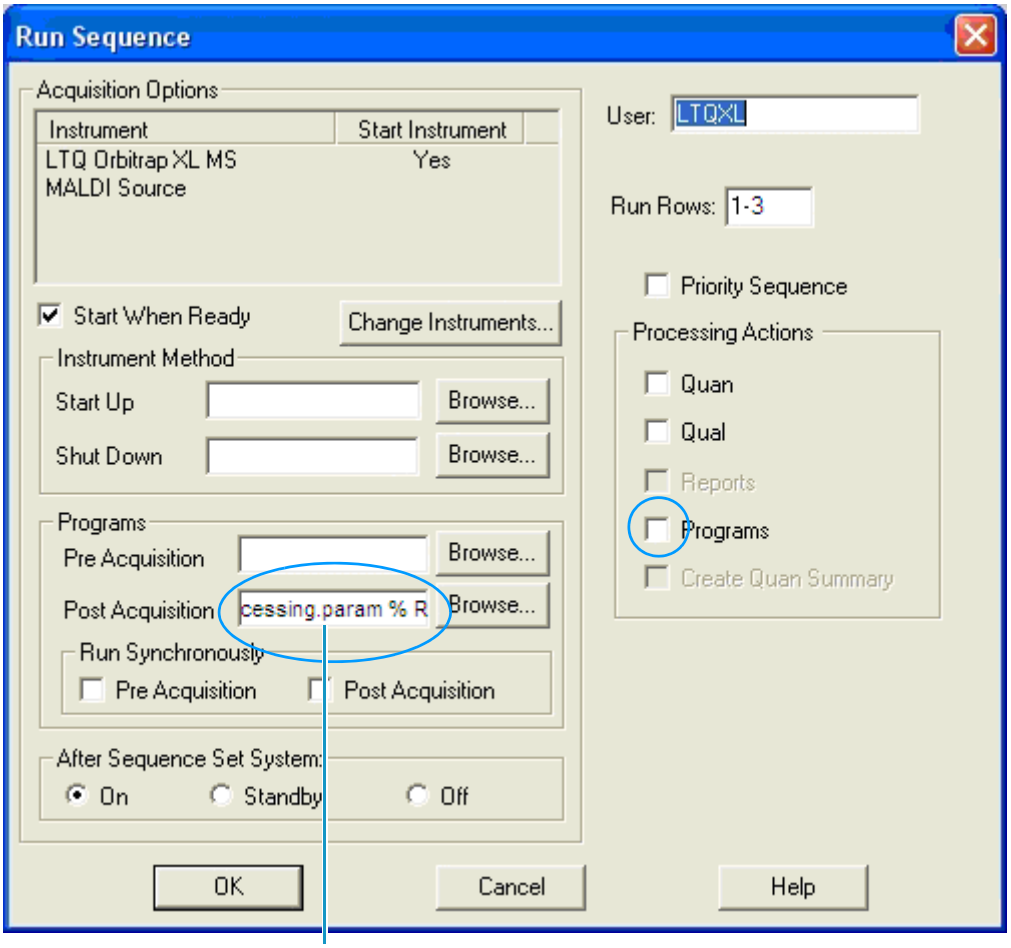
	File Name	Path	Inst Meth	Position	Proc Meth	Sample Type
1	BatchSample1	C:\Xcalibur\Data	C:\Xcalibur\methods\top50plu	H6		Unknown
2	BatchSample2	C:\Xcalibur\Data	C:\Xcalibur\methods\top50plu	H7		Unknown
3	BatchSample3	C:\Xcalibur\Data	C:\Xcalibur\methods\top50plu	H8		Unknown
*						

❖ To perform batch processing by using the Run Sequence dialog box

1. To start the sequence, click the **Run Sequence** icon,  .
2. In the Run Sequence dialog box, enter the following in the Post Acquisition box:

   C:\Program Files\Thermo\Discoverer\System\Release\discovererdaemon.exe -p C:\Xcalibur\methods\BatchProcessing.param %R

   Figure 94. Using the Proteome Discoverer Daemon application in the Run Sequence dialog box

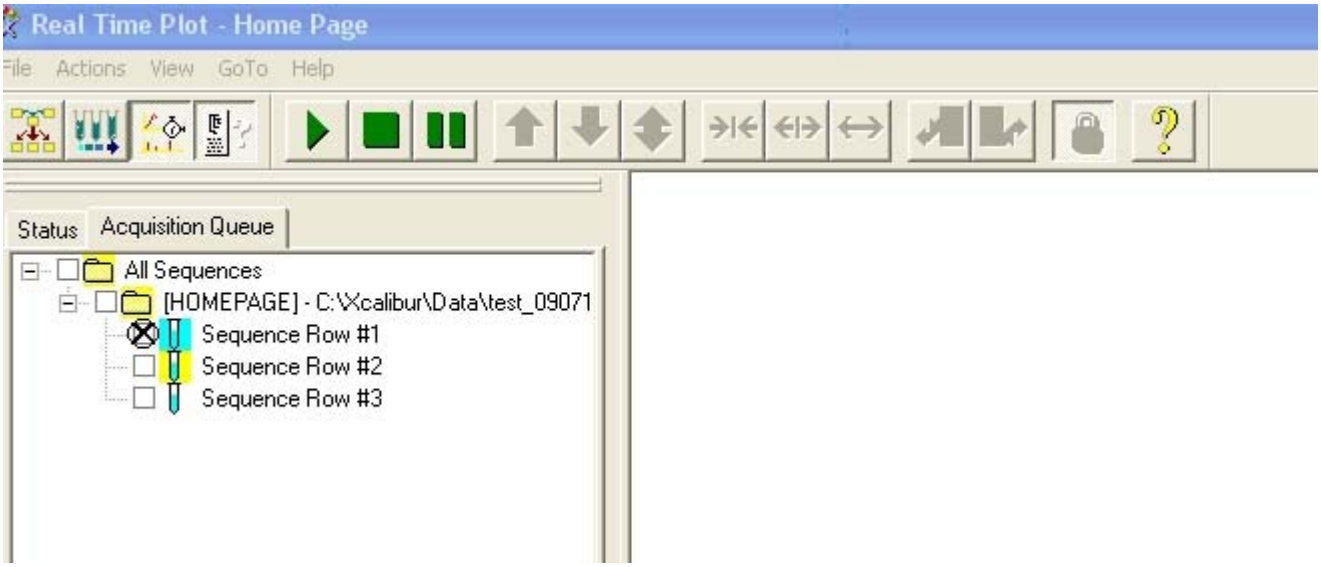


3. If the **Programs** check box in the Processing Actions area on the right is selected, clear it.

The Xcalibur data system sends the acquired raw data files synchronously to the Proteome Discoverer application, as shown in Figure 95.

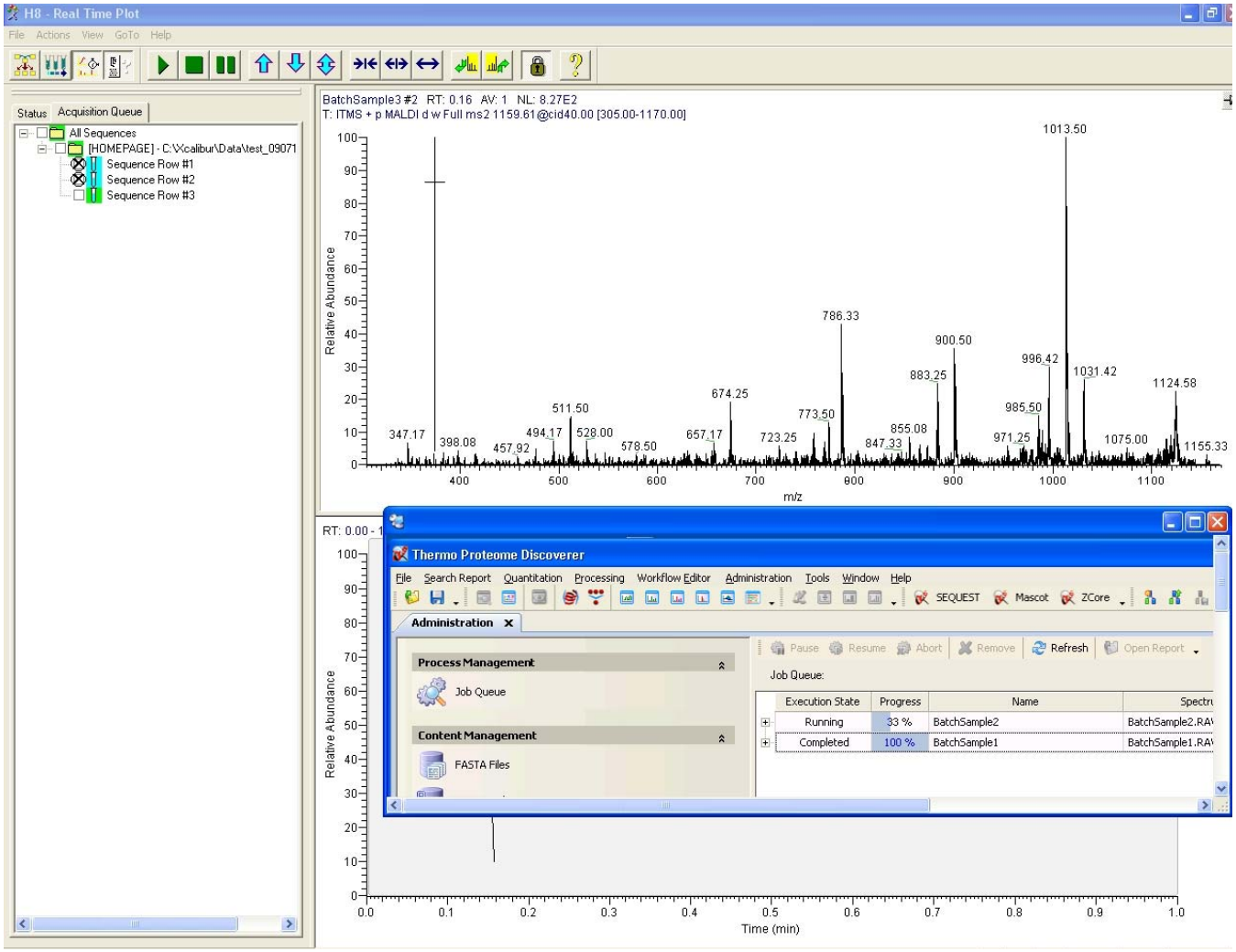
Note Only the Xcalibur 2.0.7 data system sends the acquired raw data to the Proteome Discoverer application. This functionality is not available in version 2.1.0.

Figure 95. Sending the raw data files synchronously to the Proteome Discoverer application after the first sample is finished



The Proteome Discoverer application synchronously processes the raw files on the remote host, as shown in Figure 96.

Figure 96. Processing the raw files synchronously on the remote host



In this example, the Proteome Discoverer application processes all three raw data files and places them in the Public Files directory on the computer running the Proteome Discoverer application, as shown in Figure 97 and Figure 98.

Figure 97. Completed data processing

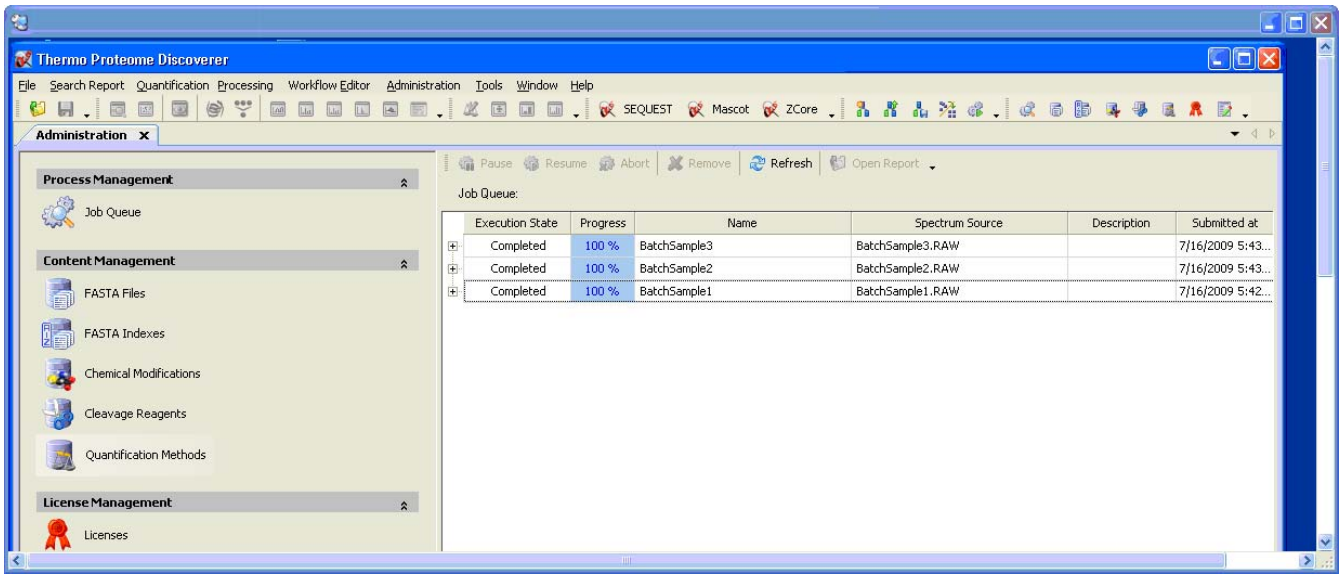
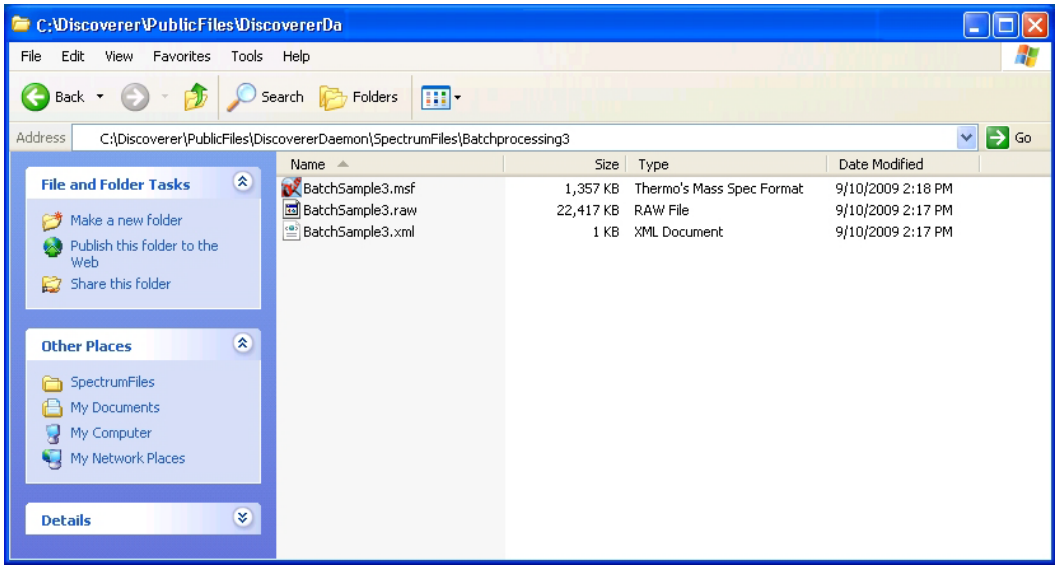


Figure 98. Storing the data in the Public Files directory



Processing MudPIT Samples by Using a Processing Method

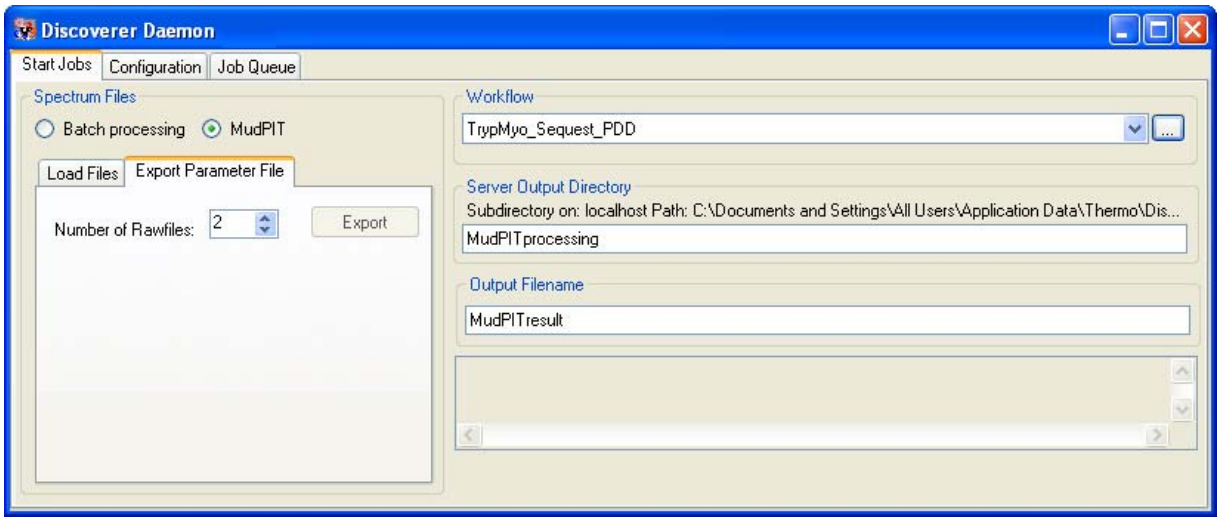
You can process MudPIT samples by using the Quantification Method Editor.

❖ To process MudPIT samples

1. Start the Proteome Discoverer Daemon application and export a parameter file for MudPIT processing. For information about exporting a parameter file, see "Creating a Parameter File for Use with the Xcalibur Data System" on page 125.

   Figure 99 shows how to configure the Export Parameter File page in the Proteome Discoverer Daemon application to export a parameter file. In the following example, the parameter file is saved in C:\Xcalibur\methods.

Figure 99. Selecting MudPIT processing in the Start Jobs page



   This example features two MudPIT samples, and each one is composed of two raw data files (for a total of four raw data files).

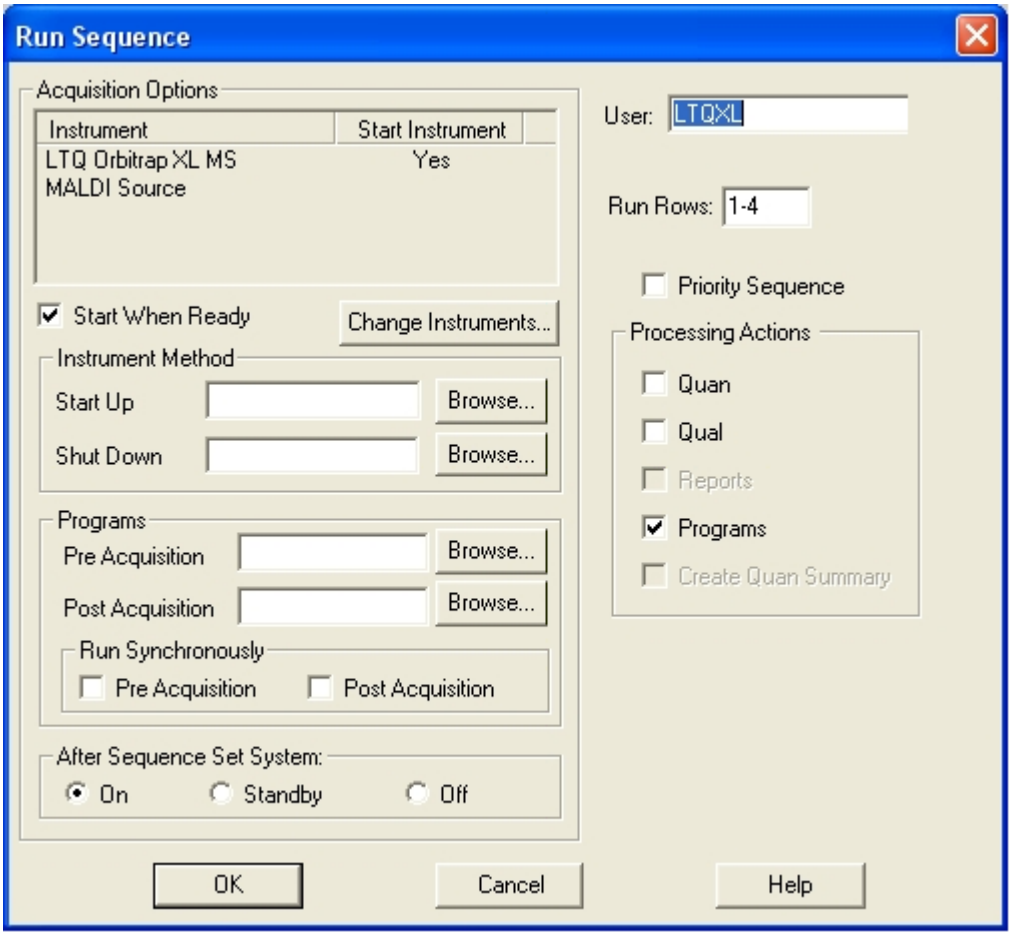
2. Define a processing method (see "Batch Processing Using a Processing Method" on page 134) using the parameter file exported in step 1, and select the method as the processing method in the Proc Meth column, as shown in Figure 100.

Figure 100. Sequence used for MudPIT processing

	File Name	Path	Inst Meth	Position	Proc Meth	Sample Type
1	MudPITSample_1	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	H6	C:\Xcalibur\methods\PDD_mudpitprocessing	Unknown
2	MudPITSample_2	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	H7	C:\Xcalibur\methods\PDD_mudpitprocessing	Unknown
3	MudPITSample_3	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	H8	C:\Xcalibur\methods\PDD_mudpitprocessing	Unknown
4	MudPITSample_4	C:\Xcalibur\Data	C:\Xcalibur\methods\top50pl	H9	C:\Xcalibur\methods\PDD_mudpitprocessing	Unknown
*						

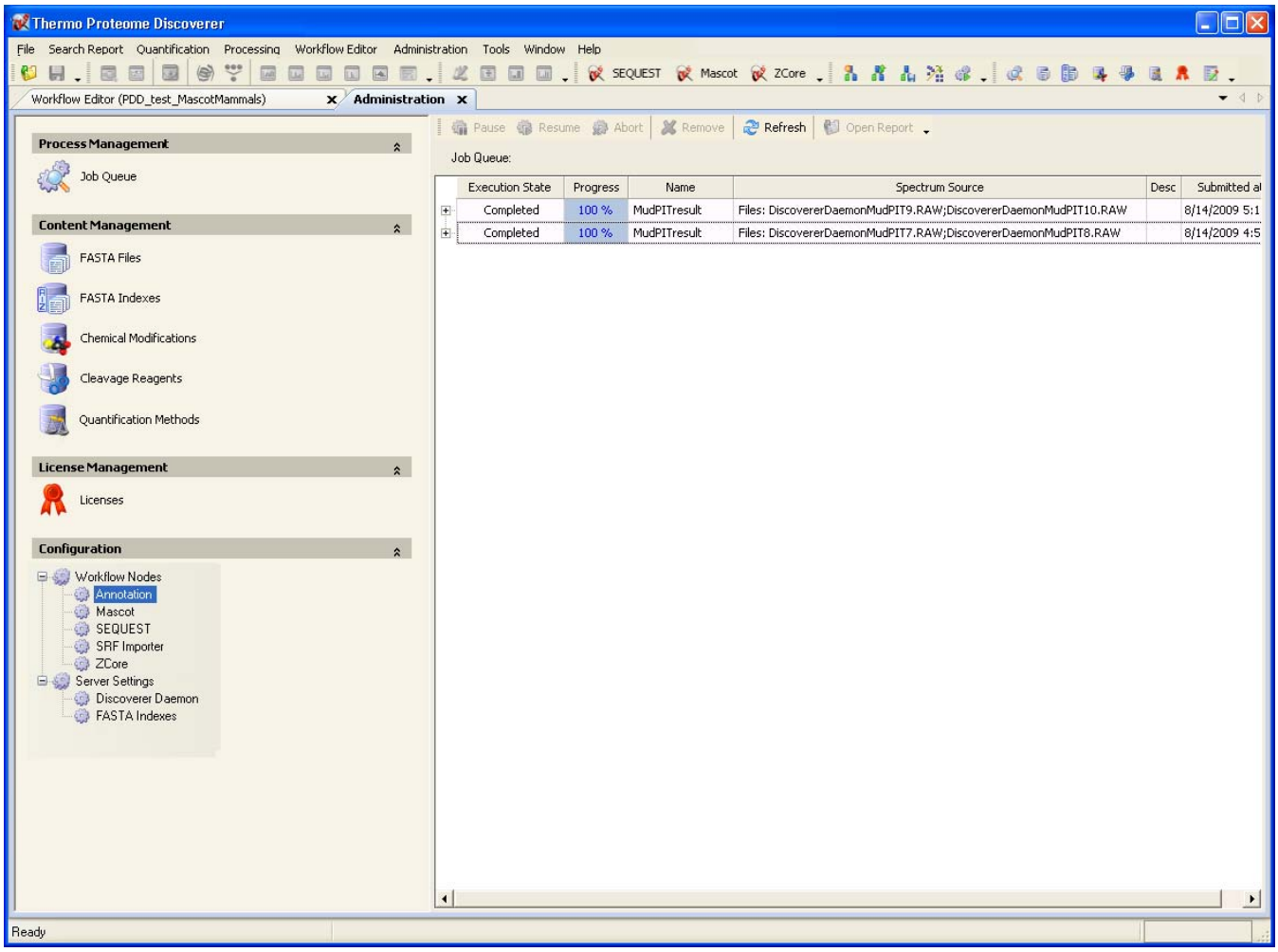
3. Start processing the MudPIT samples in the Run Sequence dialog box, as shown in Figure 101.

Figure 101. Starting the processing of the MudPIT samples



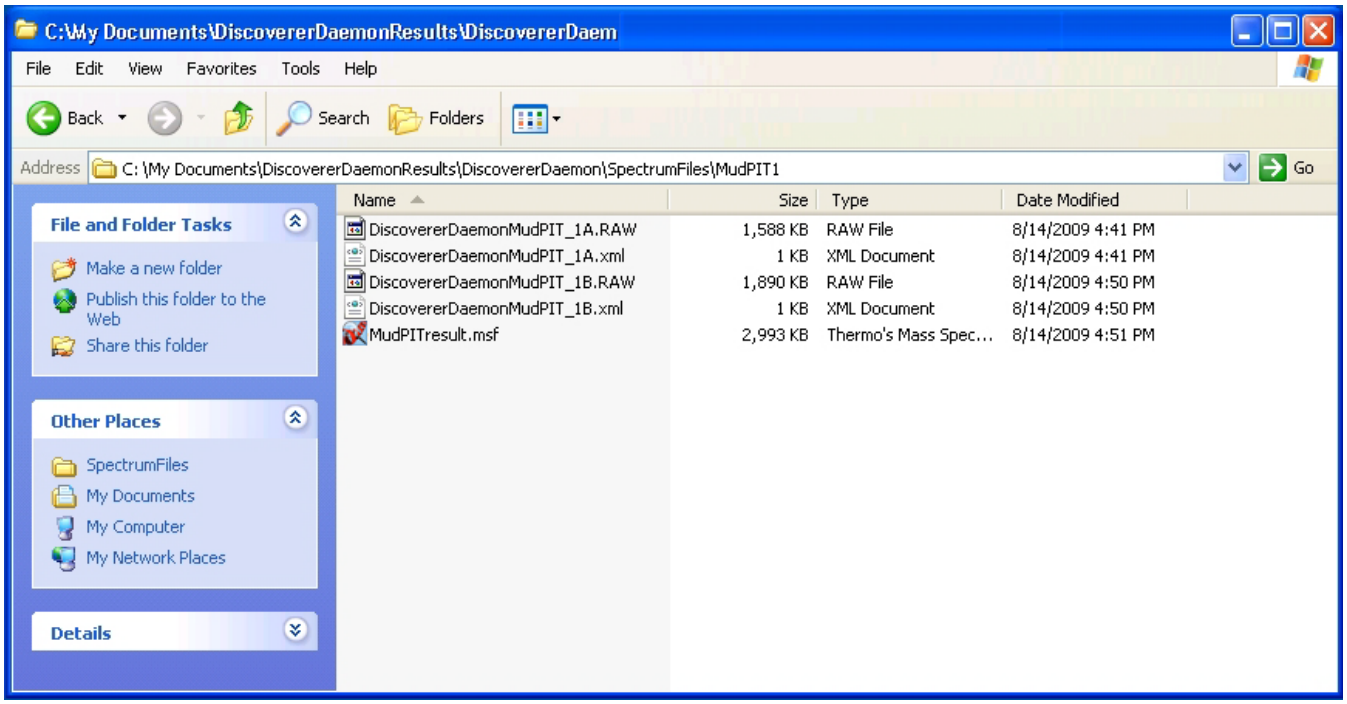
The Proteome Discoverer application processes the two samples as MudPIT, as shown in Figure 102.

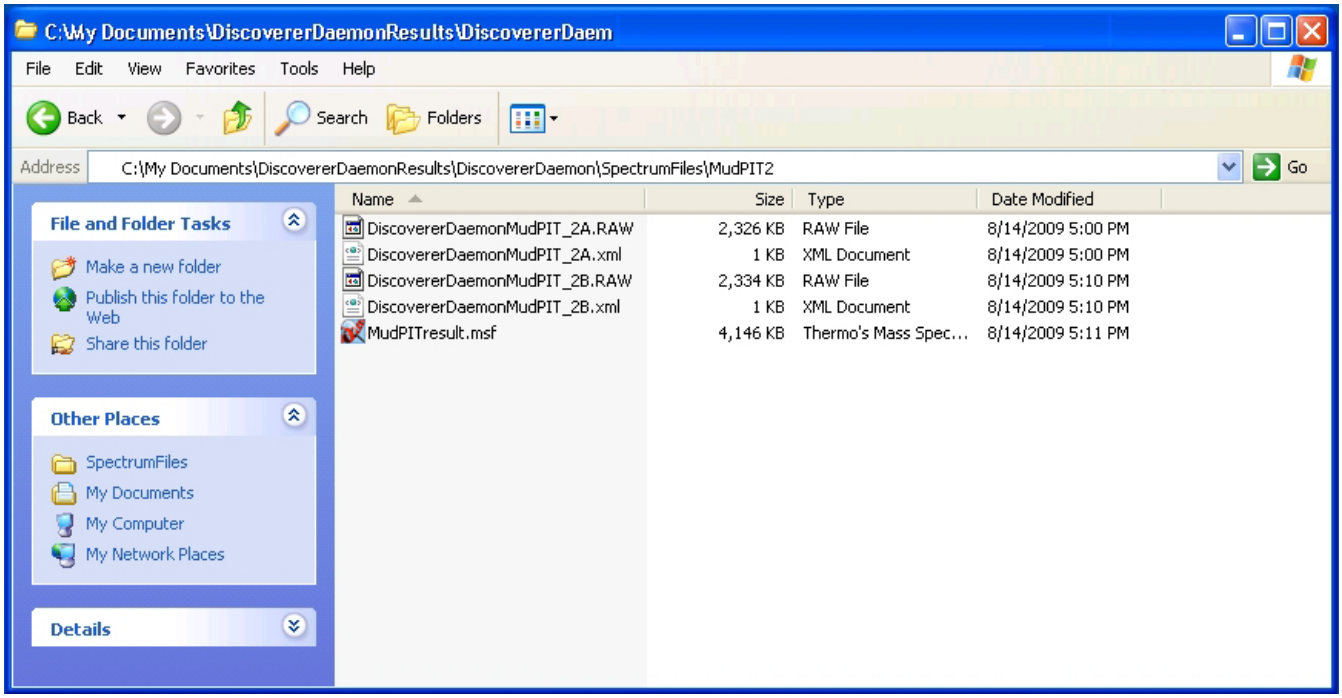
Figure 102. Processing two MudPIT samples in the Proteome Discoverer application



The Proteome Discoverer application saves the data in the two MudPIT samples in two directories, each one containing the raw data files of one MudPIT sample (in this example, two raw data files), as shown in Figure 103.

Figure 103. Saving the raw data files of each MudPIT group in two directories





MudPIT Processing Using the Run Sequence Dialog Box

Running MudPIT samples using the Run Sequence dialog box is similar to the batch processing described in "Batch Processing by Using a Post-Acquisition Method" on page 144. Replace the batchprocessing.param file with a parameter file for MudPIT.

You can use the Proteome Discoverer Daemon application to export raw files to MGF, MZDATA, DTA, MZXML, and MZML files. To export files, use a workflow that includes the Spectrum Files, Spectrum Selector, and Spectrum Exporter nodes. Set the appropriate file type in the Spectrum Exporter node. In batch processing, the Proteome Discoverer Daemon application exports all the raw files with the file name of the spectrum.

Running the Proteome Discoverer Daemon Application on the Command Line

You can run the Proteome Discoverer Daemon application on the command line or in an interface window.

❖ To run the Proteome Discoverer Daemon application on the command line

1. Open a command shell and use the cd command to move to **Program Files > Thermo > Discoverer > System > Release**.
2. Type **DiscovererDaemon** and any of the following options on the command line:

```
DiscovererDaemon
    [-e foldername FileCount Workflow ParameterAssignment]
    [-c foldername]
    [-a foldername SpectrumFile]
    [-h]
    [-l serverName userName]
    [-r outputFilename]
    [-p parameterFile rawFile]
    [-f foldername]
```

Syntax

The Discoverer Daemon command-line syntax includes the following parameters:

- `[-e foldername FileCount Workflow ParameterAssignment]`

  Executes the workflow on the server using these specified parameters:

  – *foldername*: Specifies the location where the raw files are stored. You can give it any name, for example, RawFiles or Fractions.

  - *`FileCount`*: Specifies the number of spectrum files that must be included before the workflow is executed. This parameter is intended to be used with MudPIT experiments and acquisition on several machines. If the workflow should be executed regardless of the number of files contained in the file collection, use ANY instead of a number.
  - *`Workflow`*: Specifies the name of the template file containing the workflow in .xml format. You must have created this workflow template file in the Proteome Discoverer application by choosing Workflow Editor > Export Workflow to XML.
  - *`ParameterAssignment`*: Specifies the name and value of a parameter in the format of *parameter=value*. Some examples follow.

    This example sets the FASTA database for any node to equine.fasta:

    ```
    FastaDatabase=equine.fasta
    ```

    The next example sets the FASTA database for all ZCore nodes to equine.fasta:

    ```
    ZCore.FastaDatabse=equine.fasta
    ```

    The last example sets the FASTA database for ZCore nodes having 4 as the processing node number to equine.fasta. It is equivalent to [4].FastaDatabase=equine.fasta because the processing node numbers are unique.

    ```
    ZCore[4].FastaDatabase=equine.fasta
    ```

- `[-c` *`foldername`*`]`
  - Remote server: Creates a user-named folder in the PublicFiles folder on the server where you store output files. The PublicFiles folder is the default file in the Current File Directory box in the view displayed in the Proteome Discoverer application when you select Administration > Configuration > Server Settings > Discoverer Daemon. The -c option automatically appends the date and, if the directory already exists, an incremental index number to the name.

    You can only create a folder in the directory configured in the view opened by the Administration > Configuration > Server Settings > Discoverer Daemon command on the remote server. If you attempt to create a file other than in the PublicFiles folder in the Current File Directory box, Discoverer Daemon issues a message informing you that the Proteome Discoverer application will apply the change the next time that you start it.

    This option performs the same function as the -f *foldername* option, except that you can use the name of the folder more than once. When you use the name more than once, the Proteome Discoverer application appends the date and an incremental index number to the name.
  - Local server: Does nothing.

- `[-a foldername SpectrumFile]`
  - Remote server: Uploads the spectrum file to the location specified on the configured server. *`SpectrumFile`* is the name of the spectrum file.
  - Local server: Does nothing.
- `[-h]`: Lists the options available with the Thermo.Magellan.DiscovererDaemon command.
- `[-l serverName userName]`: Connects Discover Daemon to the specified local or remote host machine.
  - *`serverName`*: Specifies the name of the local or remote host.
  - *`userName`*: Specifies the name to log on.
- `[-r outputFilename]`: Specifies the name of the output file. You must use this option with the `-e` option, as in this example:

  ```
  DiscovererDaemon -e sfcid any mascot3.xml -r silac1noMT_AS4DE.msf.
  ```
- `[-p parameterFile rawFile]`: Processes the specified raw data file with all the parameters given in the parameter file, including the connection to the server.
  - *`parameterFile`*: Specifies the name of the parameter file.
  - *`rawFile`*: Specifies the name of the raw file.

  In the following example of the -p syntax, the Proteome Discoverer Daemon application processes the 9mix_LysC_monolith.raw file with the parameters given in the parameter file called c:\Xcalibur\methods\batchprocessing.param.

  ```
  DiscovererDaemon -p C:\Xcalibur\methods\batchprocessing.param
  9mix_LysC_monolith.raw
  ```
- `[-f foldername]`: On a remote server, this option creates a user-named folder in the PublicFiles folder of the server where the local version of the raw file and the result files are stored. If the directory already exists, the Proteome Discoverer Daemon application issues an error message, and the process returns with exit code -1 (standard exit code 0).

  If you attempt to create a file other than in the PublicFiles folder in the Current File Directory box, Discoverer Daemon issues a message informing you that the Proteome Discoverer application will apply the change the next time that you start it.

  This option performs the same function as the -c *foldername* option, except that you cannot use the name of the folder more than once.

  On a local server, this option does nothing.

Examples

The following are some examples of the Proteome Discoverer Daemon command-line syntax.

This example constructs the spectrum file collection called Rawfiles, adds the TrypMyo.raw file to the collection, and executes the SequestEquine workflow using the raw file in the Rawfiles directory:

```
DiscovererDaemon -c Rawfiles -a Rawfiles
c:\Rawfiles\TrypMyo.raw -e Rawfiles ANY c:\Workflows\SequestEquine.xml
```

In the following example, the Proteome Discoverer Daemon application evaluates several fractions in a single workflow:

```
DiscovererDaemon -c Fractions
DiscovererDaemon -a Fractions c:\rawfiles\fraction1.raw
DiscovererDaemon -a Fractions c:\rawfiles\fraction2.raw
DiscovererDaemon -a Fractions c:\rawfiles\fractionN.raw
DiscovererDaemon -e Fractions ANY c:\wfs\fractions.xml
```

The next example demonstrates that you can start several workflows with one invocation of the Proteome Discoverer Daemon application.

```
DiscovererDaemon
-c RawFile
-a RawFile c:\Rawfiles\TrypMyo.raw
-e RawFile ANY c:\wfs\SequestEquine.xml
-c RawFile
-a RawFile c:\Rawfiles\BSADigest.raw
-e RawFile ANY c:\Workflows\SequestEquine.xml
```

The following example runs the Proteome Discoverer Daemon application on a remote host called protlab2, uploads the iTRA_BSA_3ITMS2_3HCD.raw spectrum file to the server, executes the workflow in c:\Workflows\MascotEcoli.xml, and deletes the uploaded raw file from the server:

```
DiscovererDaemon -l protlab2 leo_davinci -c sfcid -a sfcid
iTRA_BSA_3ITMS2_3HCD.raw -e sfcid any c:\Workflows\MascotEcoli.xml
```

The following sequence of commands submits multiple raw files for processing on a remote server:

```
DiscovererDaemon.exe -c AllTrypMyo
DiscovererDaemon.exe -a AllTrypMyo_020110303 C:\DaemonTest\mudpit4\Tryp_Myo.raw
DiscovererDaemon.exe -a AllTrypMyo_020110303 C:\DaemonTest\mudpit4\Tryp_Myo_1.raw
DiscovererDaemon.exe -a AllTrypMyo_020110303 C:\DaemonTest\mudpit4\Tryp_Myo_2.raw
DiscovererDaemon.exe -e AllTrypMyo_020110303 3 C:\DaemonTest\mudpit4\wf_sequest.xml
```

The next sequence of commands submits multiple raw files for processing on a local server:

```
DiscovererDaemon.exe -a AllTrypMyo C:\DaemonTest\mudpit4\Tryp_Myo.raw
DiscovererDaemon.exe -a AllTrypMyo C:\DaemonTest\mudpit4\Tryp_Myo_1.raw
DiscovererDaemon.exe -a AllTrypMyo C:\DaemonTest\mudpit4\Tryp_Myo_2.raw
DiscovererDaemon.exe -e AllTrypMyo 3 C:\DaemonTest\mudpit4\wf_sequest.xml
```

The Discoverer Daemon appends a time stamp to each file when it processes the files on a remote server.

Searching for Data

This chapter describes the features that you can use when searching for and analyzing data in the Proteome Discoverer application.

Contents

- Using FASTA Databases
- Chemical Modifications
- Using the Qual Browser Application
- Customizing Cleavage Reagents

Using FASTA Databases

You can use the FASTA database utilities to add, delete, and find protein references and sequences. You can also extract information from an existing FASTA file, place it into a new FASTA file, and compile it for availability in the Proteome Discoverer application.

For more information about FASTA databases, see "FASTA Reference" on page 705.

Displaying FASTA Files

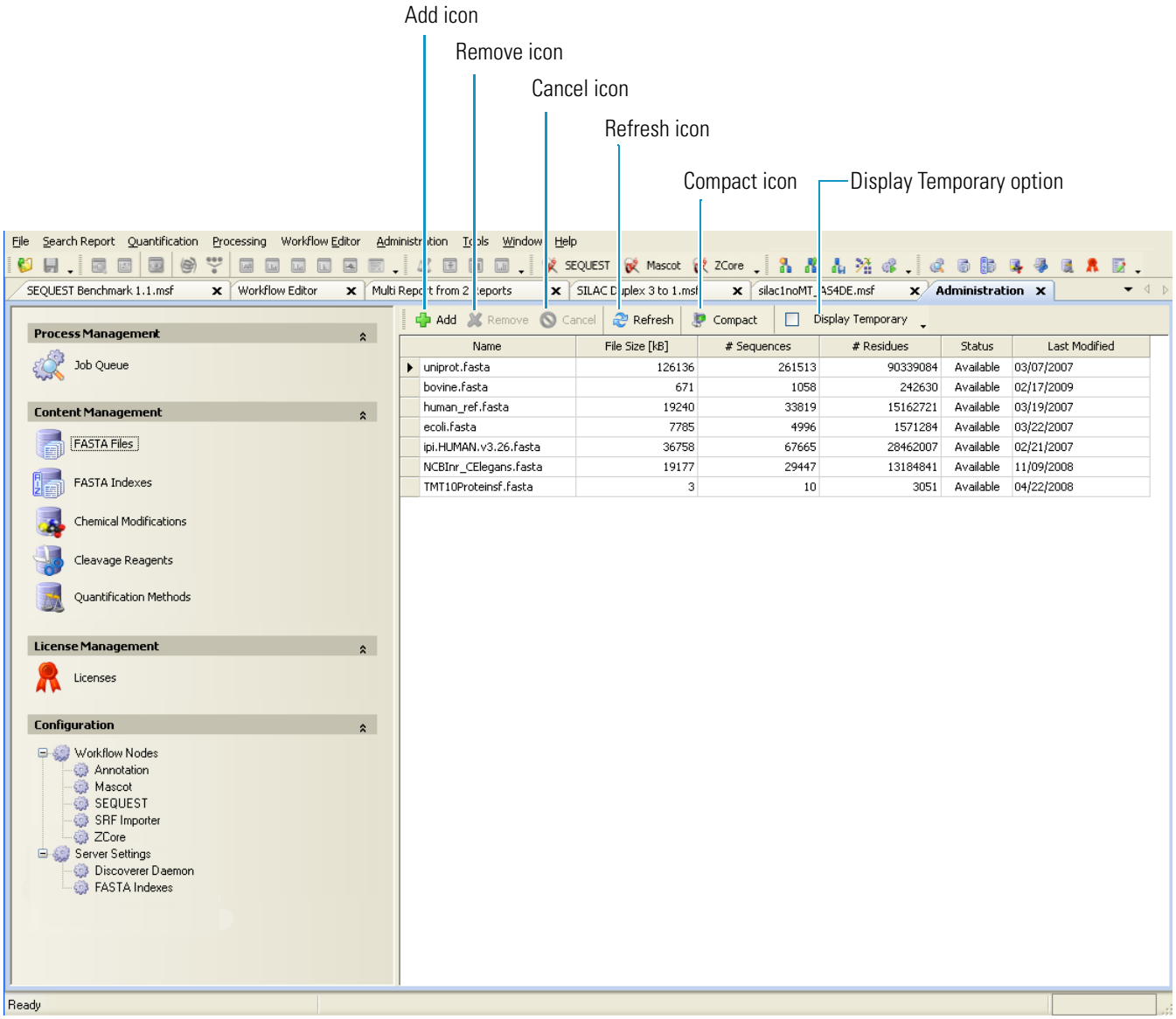
You can list all the FASTA files that you have downloaded from other sources onto your hard drive and registered.

❖ **To list the available FASTA files**

- Choose **Administration > Maintain FASTA Files** or click the **Maintain FASTA Files** icon,  , either in the toolbar or on the Administration page.

The FASTA files view shown in Figure 104 appears. It lists all the FASTA files that you have downloaded from other sources and registered. It displays the processed FASTA file properties, such as the file name, file size, and the number of proteins stored. The Proteome Discoverer application analyzes each protein entry to determine if the FASTA file meets the application requirements for use in a spectra search. It processes the FASTA file and makes it available for use.

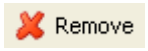
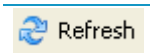
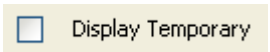
Figure 104. FASTA files view



FASTA Files View Parameters

Table 26 describes the options and columns in the FASTA files view in the Proteome Discoverer application.

Table 26. Options and columns in the FASTA files view

Parameter	Description
	Activates the Open dialog box, so you can choose the FASTA database to import.
	Deletes a FASTA database from the FASTA files view.
	Cancels the addition or removal of a FASTA file.
	Redisplays the view on the screen.
	Releases the storage space previously occupied by proteins that were imported from FASTA files and inserted during a Mascot search but subsequently deleted.
	Displays FASTA files that contain the proteins found by a Mascot search. The Proteome Discoverer application temporarily imports these FASTA files, which are not available for SEQUEST and ZCore searches.
Name	Displays the name of the FASTA file.
Size [kB]	Displays the current size of the FASTA file.
#Sequences	Displays the number of sequences found in the FASTA file during processing.
#Residues	Displays the number of amino acids found in the FASTA file during processing.
Status	Displays the current status of the FASTA file: • Imported: Indicates that the FASTA file has been downloaded from a source and registered. • Available: Indicates that the FASTA file is available for SEQUEST and ZCore searches. • Processing: Indicates that the FASTA file is in the process of being registered.
Last Modified	Displays the date when the FASTA file was last modified or created.

Adding FASTA Files

You must add a FASTA file to the Proteome Discoverer application before you can conduct a search with the SEQUEST or ZCore search engine.

❖ To add a FASTA file

1. Choose **Administration > Maintain FASTA Files** or click the **Maintain FASTA Files** icon,  .

   The Administration page appears with the FASTA files view, shown in Figure 104 on page 160.

2. Click  .

3. In the Open dialog box that appears, browse for and select the FASTA file that you want to process, and then click **Open**.

   The FASTA file that you selected appears as a job in the job queue. To cancel the addition of this file, click  .

4. To add another FASTA file, wait until the Execution State column indicates that the addition of the FASTA file is completed, click **FASTA Files** in the left pane of the Administration page under Content Management, and then click **Add** to add the next file.

The amount of time that it takes to process a FASTA file depends on the file size. When a FASTA file finishes processing, the Status column displays the Available status. The FASTA file is now available to use for a protein or peptide search with the Proteome Discoverer application.

Deleting FASTA Files

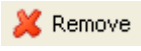
You can delete a FASTA file from the application.

❖ To delete a FASTA file

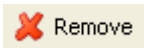
1. Choose **Administration > Maintain FASTA file**.

   The Administration page appears with the FASTA files view, shown in Figure 104 on page 160.

2. Click  at the beginning of a row to select the row.

3. Click  .

4. In the Remove FASTA databases dialog box, click **OK**.

The FASTA file that you selected appears as a job in the job queue. After you start the deletion of the file, you cannot cancel the deletion. You can remove the completed job from the job queue by clicking  and then clicking **OK** in the Delete Jobs dialog box.

Compressing a Protein Database

A protein database contains the proteins of imported FASTA files. It also contains proteins found during a Mascot search that are inserted into the database. When you remove a FASTA file from the database by using the FASTA file manager, it automatically deletes protein entries but does not make the storage space available. Although following this next procedure can explicitly make the storage space available, it can be time-consuming for large databases.

❖ To compress a protein database

1. Choose **Administration > Maintain FASTA file**.

   The Administration page appears with the FASTA files view.

2. Click 

.

   A message informs you that compressing the protein database can take a long time.

3. To continue with the database compression, click **OK** in the message box.

   A job starts and appears in the job queue. Before you start the job, you can remove it if necessary. However, you cannot cancel the job, and it will restart automatically if you shut down the Proteome Discoverer application during job execution.

Displaying Temporary FASTA Files

The Proteome Discoverer application temporarily imports FASTA files that contain the proteins found by a Mascot search, but these files are not available for SEQUEST and ZCore searches. You can optionally display these files in the FASTA files view.

❖ To display temporary FASTA files

1. Choose **Administration > Maintain FASTA file**.

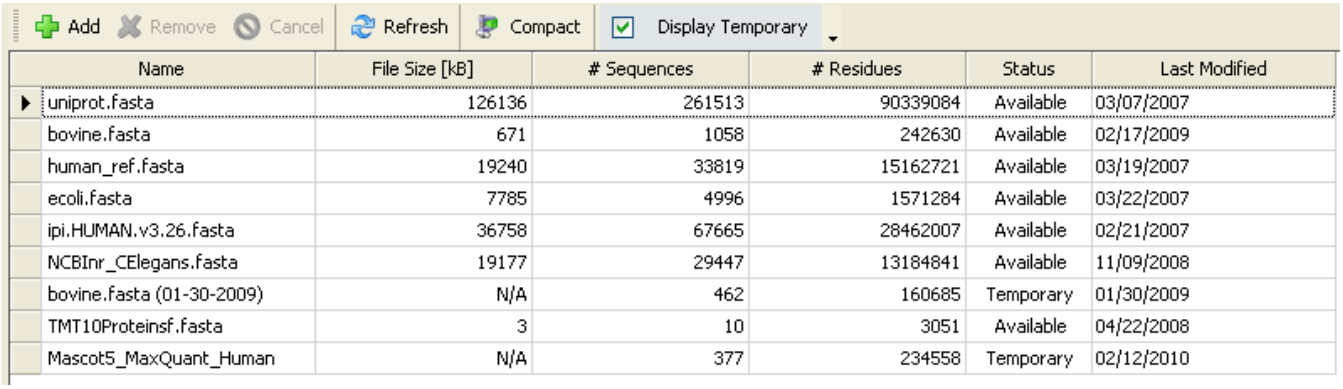
   The Administration page appears with the FASTA files view, shown in Figure 104 on page 160.

2. Select the **Display Temporary** check box, 

.

   You now see any temporary FASTA files; for example, Figure 105 shows Temporary for two files in the Status column.

Figure 105. Displaying temporary FASTA files



Name	File Size [kB]	# Sequences	# Residues	Status	Last Modified
uniprot.fasta	126136	261513	90339084	Available	03/07/2007
bovine.fasta	671	1058	242630	Available	02/17/2009
human_ref.fasta	19240	33819	15162721	Available	03/19/2007
ecoli.fasta	7785	4996	1571284	Available	03/22/2007
ipi.HUMAN.v3.26.fasta	36758	67665	28462007	Available	02/21/2007
NCBInr_CElegans.fasta	19177	29447	13184841	Available	11/09/2008
bovine.fasta (01-30-2009)	N/A	462	160685	Temporary	01/30/2009
TMT10Proteinsf.fasta	3	10	3051	Available	04/22/2008
Mascot5_MaxQuant_Human	N/A	377	234558	Temporary	02/12/2010

Adding a Protein Sequence and Reference to a FASTA Database File

You can add a protein sequence and a protein reference to a registered FASTA database file. The protein sequence refers to the sequence of amino acids that constitute the protein, and the protein reference refers to the name or reference of the protein.

❖ To add a protein sequence and reference

1. Choose **Tools > FASTA Database Utilities**.
2. In the FASTA Database Utilities dialog box, click the **Add Protein References** tab.

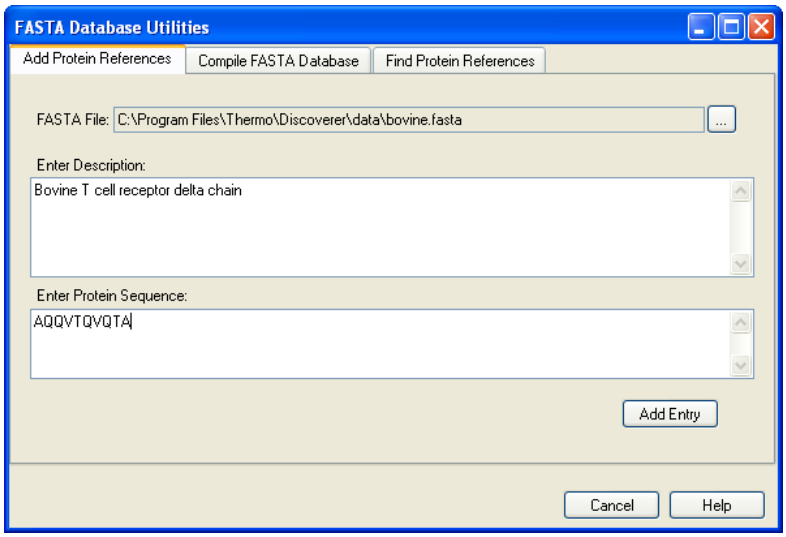
   The Add Protein References page of the dialog box appears.

3. Click the browse button (...) next to the FASTA File box.
4. In the Save/Add to FASTA File dialog box, select the FASTA database that you want to add the protein sequence and reference to, and click **Save**.

5. In the Enter Description box of the FASTA Database Utilities dialog box, type a description of the protein sequence that you are adding.
6. In the Enter Protein Sequence box, type the protein sequence that you want to add to the FASTA database.

   The Add Protein References page should resemble the illustration in Figure 106.

   Figure 106. Add Protein References page of the FASTA Database Utilities dialog box



7. Click **Add Entry** to add the protein sequence.

Finding Protein Sequences and References

You can find a protein sequence or reference in an existing FASTA database file.

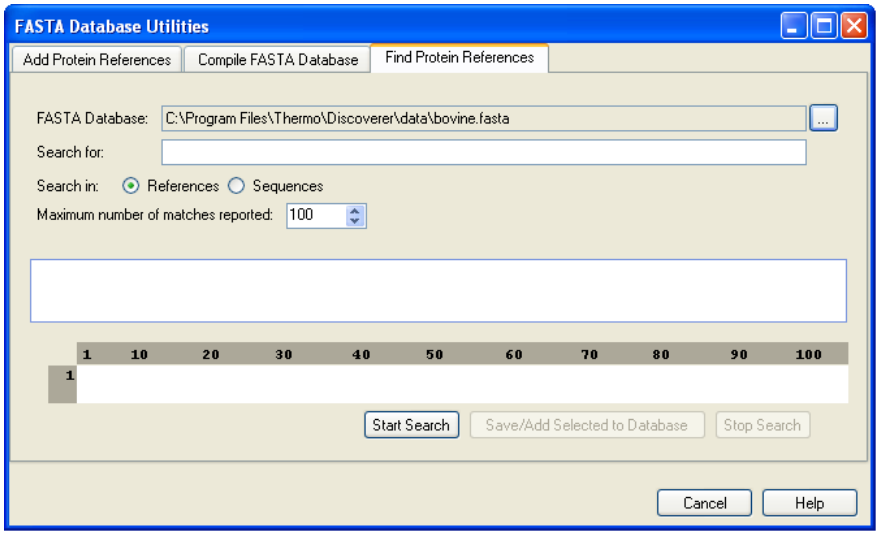
- To find a protein sequence or reference
- To filter a protein reference search
- To refine a filtered protein reference search
- To delete conditions in filtered protein reference searches

❖ To find a protein sequence or reference

1. Choose **Tools > FASTA Database Utilities**.
2. In the FASTA Database Utilities dialog box, click the **Find Protein References** tab.

   The Find Protein References page appears, as shown in Figure 107.

Figure 107. Find Protein References page of the FASTA Database Utilities dialog box

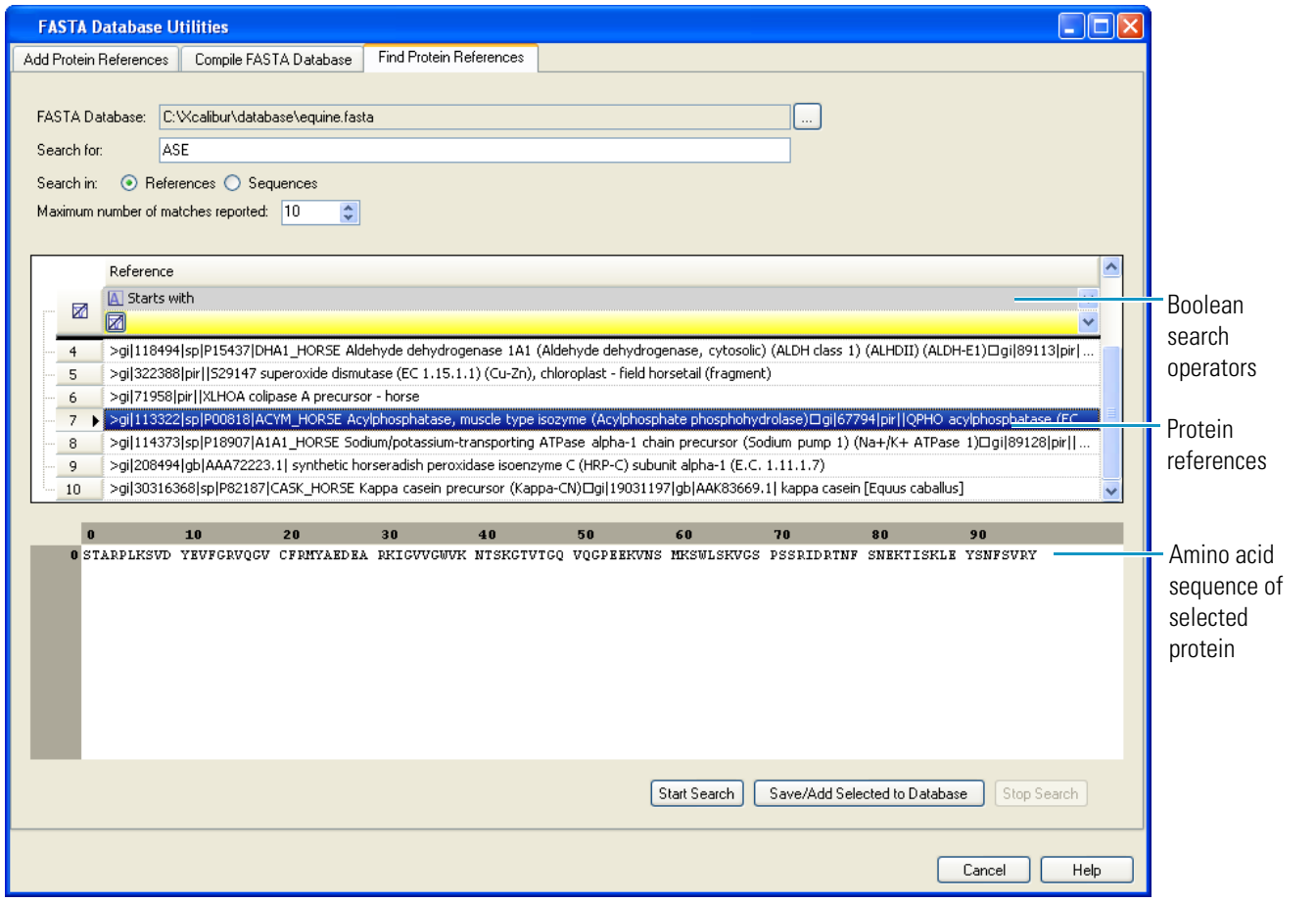


3. Click the browse button (...) next to the FASTA Database box to locate the FASTA file of interest.
4. In the Please Select a FASTA Database dialog box, select the FASTA file, and click **Open**.
5. In the Search For box of the Find Protein References page, type an amino acid sequence or a protein reference search string.
6. In the Search In box area, select the **References** option (the default) to search for a reference or the **Sequences** option to search for a sequence.

   You can further refine the results by using filters either before or after you run the search. For instructions on filtering, see "To filter a protein reference search" on page 167.
7. In the Maximum Number of Matches Reported box, select the maximum number of references or sequences to report.
8. Click **Start Search**.

   Results appear if the search parameters match the data, as shown in Figure 108. Click a protein row to see the amino acid sequences that constitute that protein.
9. To suspend the search, click **Stop Search**.

Figure 108. Find Protein References page in the FASTA Database Utilities dialog box



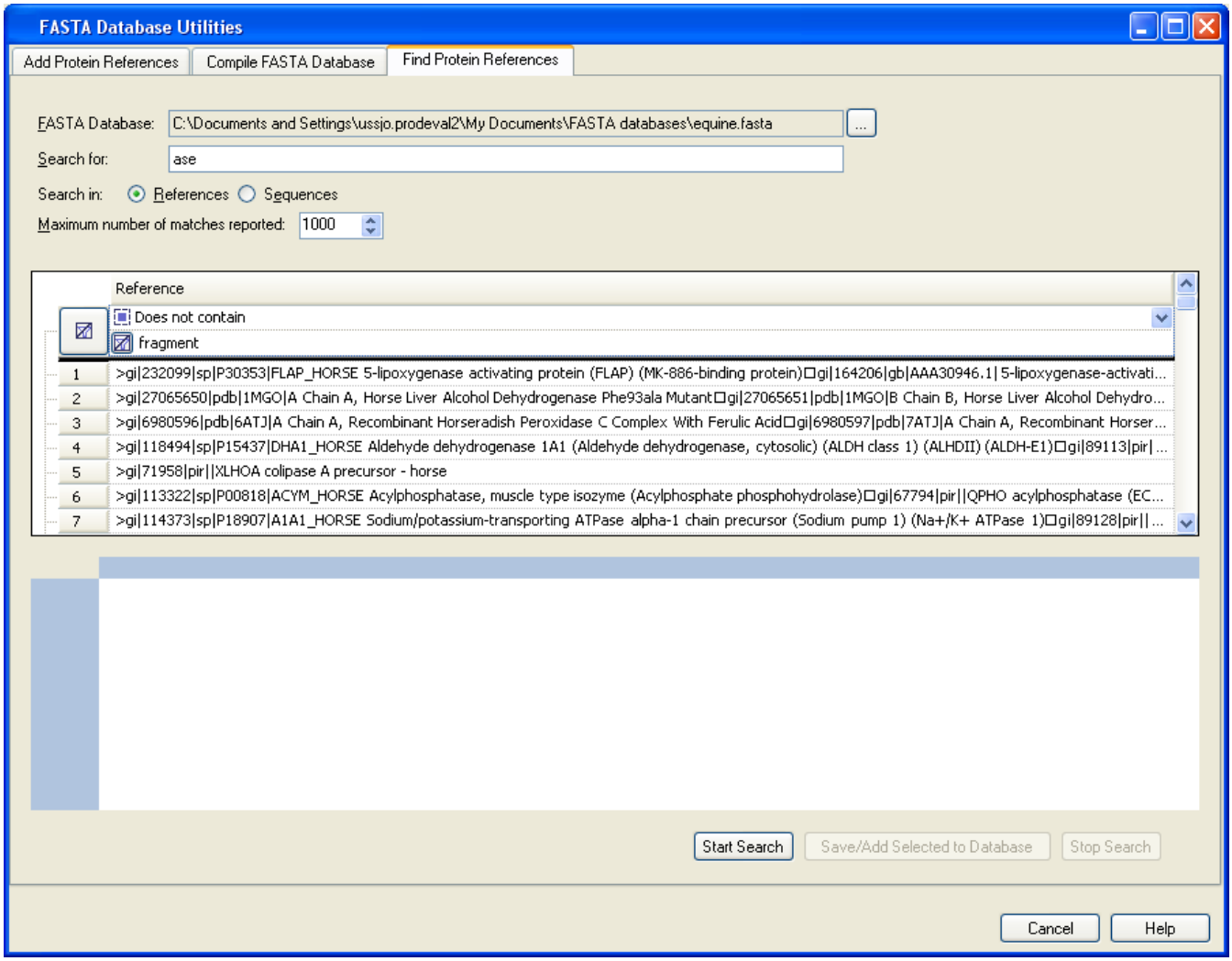
10. (Optional) To save a protein result row in another FASTA database, select the protein row, click **Save/Add Selected to Database**, select the database in the Save/Add to FASTA File dialog box, and click **Save**.

❖ To filter a protein reference search

1. On the Find Protein References page of the FASTA Database Utilities dialog box, click the line below "Reference" in the middle of the page to access a list of operators that you can use to filter the references. (The default operator is "Starts with.") For a list of all operators, see "Find Protein References Page" on page 177.
2. In the line below the operator that you selected, type the search string or condition that you want the operator to apply to.

   The example in Figure 109 filters out those protein references that contain "fragment."

Figure 109. Filtering out protein references containing “fragment”

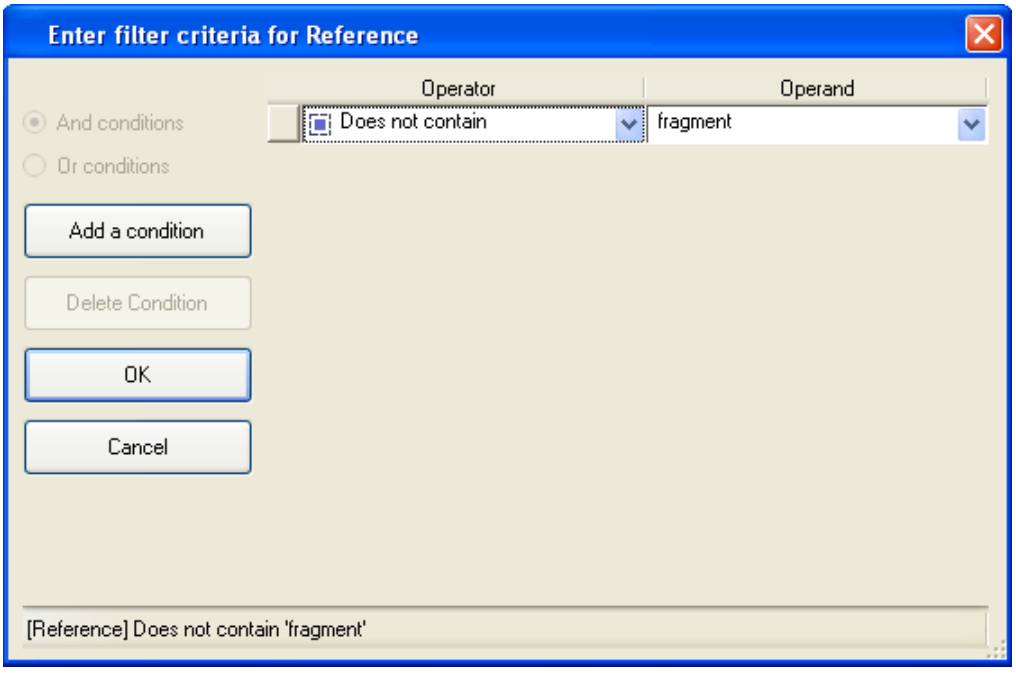


❖ To refine a filtered protein reference search

1. Select the **Custom** option from the list in the line below the operator.

   This option opens the Enter Filter Criteria for Reference dialog box, as shown in Figure 110, so you can add multiple conditions.

Figure 110. Enter filter criteria for Reference dialog box



2. Click **Add a Condition**.

   A new line appears in the Operator and Operand lists.

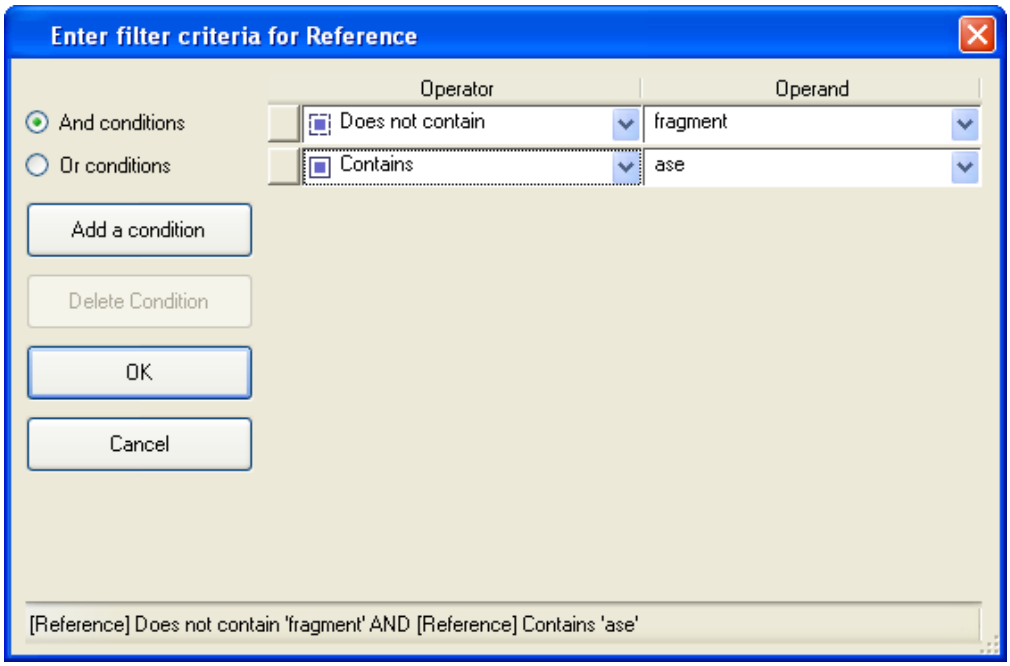
3. Select an operator from the Operator list.

4. Type an operand on the line in the Operand column.

5. Select the **And Conditions** option to indicate whether the search algorithm should search for protein references that meet both conditions, or select the **Or Conditions** option to indicate whether the search algorithm should search for protein references that meet only one of the conditions.

   Figure 111 gives an example of a search for protein references that meets both of the conditions.

Figure 111. Specifying a second condition



6. Click **OK**.

❖ **To delete conditions in filtered protein reference searches**

- To delete a condition in the Enter filter criteria for Reference dialog box, select the box to the left of the appropriate condition in the Operator column, and click **Delete Condition**.
- To delete the condition under "Reference" on the Find Proteins References page, click the **Clear Reference Filter** icon, .
- To delete all conditions in both the Enter filter criteria for Reference dialog box and the condition under "Reference" on the Find Proteins References page, click the **Clear All Filters** icon, .

Compiling a FASTA Database

You can extract information from an existing FASTA file and place it into a new FASTA file, replace an existing FASTA file, or append it to an existing FASTA file. Then you must compile the new or changed FASTA file to make it available in the Proteome Discoverer application.

❖ **To compile a FASTA database**

1. Choose **Tools > FASTA Database Utilities**.
2. In the FASTA Database Utilities dialog box, click the **Compile FASTA Database** tab.

   The Compile FASTA Database page appears.

3. In the Original box, browse for the FASTA file that you are taking the information from, or type its path and name.
4. In the Please Select a FASTA Database dialog box, click **Open**.
5. In the Target box, browse for the FASTA file that you are placing the extracted information into, or type its path and name.
6. In the Save/Add to FASTA File dialog box, select the file, verify that the file extension is .fasta, and click **Save**.
7. In the Target Database Options area, select one of the following options to indicate what you want to do with the extracted information:
   - Create/Replace: Creates a new FASTA file for storing the information or overwriting an existing FASTA file. This option is the default.
   - Append: Adds the extracted information to an existing FASTA file.
8. To disregard the case of the information to be extracted, select the **Ignore case of reference strings** check box.
9. Specify the information to be extracted:
   a. Click  above the Step 1: String(s) to Include box.

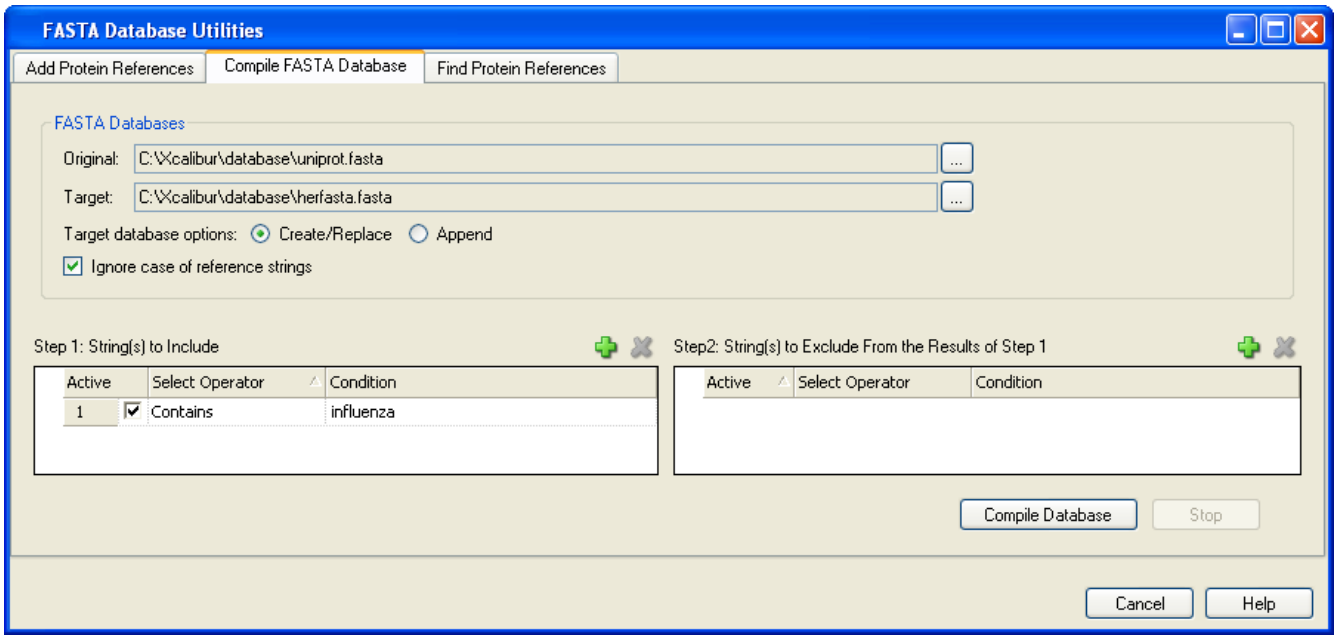
      A line enabling you to specify the first set of conditions appears in the box.

   b. Click the first line in the Select Operator column, and select the operator to apply to the information to be extracted. You can select from the following:
      - Starts With: Extracts information that begins with this string.
      - Does Not Start With: Extracts information that does not begin with this string.
      - Ends With: Extracts information that ends with this string.
      - Does Not End With: Extracts information that does not end with this string.
      - Contains: Extracts information that includes this string.
      - Does Not Contain: Extracts information that does not includes this string.

   c. Click the first line in the Condition column, and type the condition that the information must meet in order to be extracted.
   d. Repeat step a through step c to add more sets of conditions for the information to be extracted.
   e. To delete a set of conditions, in the Active column select the line that you want to delete and click  .

   The Compile FASTA Database page should now resemble the example in Figure 112.

Figure 112. Compile FASTA Database page of the FASTA Database Utilities dialog box



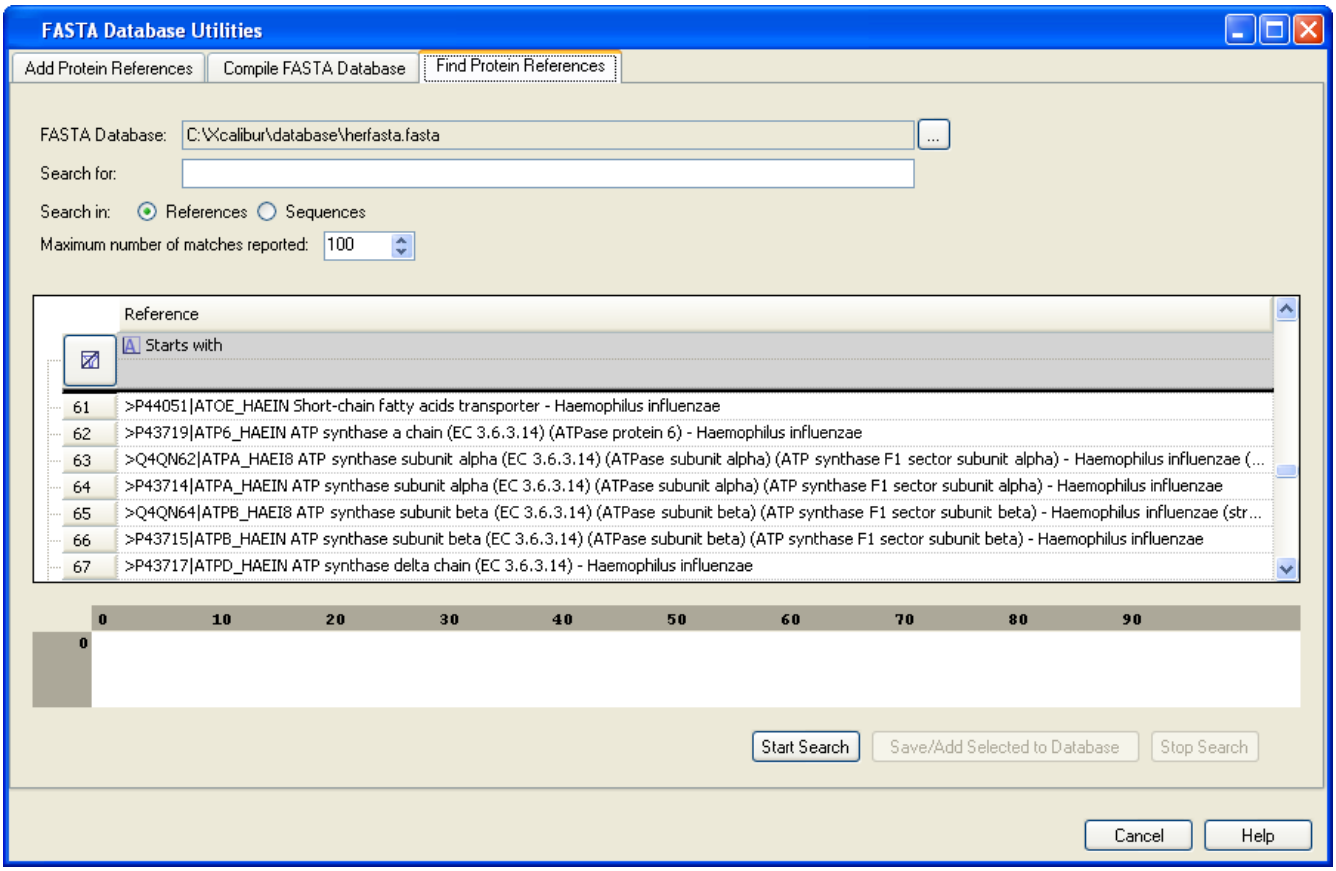
10. Click **Compile Database**.

    Click **Stop** to halt the compilation.

11. After the compilation, click **Start Search** on the Find Protein References page to view the results of the extraction, as shown in the example in Figure 113.

    You do not have to enter information into the Search For box.

Figure 113. Results of search



12. (Optional) To specify any information that you want to exclude from the extracted results, follow these steps:

    a. Click  above the Step 2: String(s) to Exclude From the Results of Step 1 box on the Compile FASTA Database page.

    A line enabling you to specify the first set of conditions now appears in the box.

    b. Click the first line in the Select Operator column, and select the operator to apply to the information from the list. You can choose from the following:

    - Starts With: Excludes information that begins with this string.
    - Does Not Start With: Excludes information that does not begin with this string.
    - Ends With: Excludes information that ends with this string.
    - Does Not End With: Excludes information that does not end with this string.
    - Contains: Excludes information that includes this string.
    - Does Not Contain: Excludes information that does not include this string.

    c. Click the first line in the Condition column, and type the condition that the information must meet in order to be excluded.

d. Repeat step a through step c to add more sets of conditions for the information that you want to exclude.

e. To delete a set of conditions, in the Active column select the line that you want to delete and click  .

13. Click **Compile Database**.

14. Click **Start Search** on the Find Protein References page to view the results of the extraction, as shown in the example in Figure 113.

    You do not have to enter information into the Search For box.

Excluding Individual Protein References and Sequences from a FASTA Database

You can exclude individual entries from a FASTA file.

❖ To exclude individual protein references and sequences from a FASTA file

1. Choose **Tools > FASTA Database Utilities**.
2. In the FASTA Database Utilities dialog box, click the **Compile FASTA Database** tab.
3. In the Original box, browse for the FASTA database that contains the protein that you want to remove, or type its path and name. In the Please Select a FASTA Database dialog box, click **Open**.
4. In the Target box, browse for the output FASTA file or type its path and name. In the Save/Add to FASTA File dialog box, select the file, verify that the file extension is .fasta, and click **Save**.
5. Select the **Ignore Case of References Strings** check box.
6. Click  above the Step 1: String(s) to Include box.

   A line enabling you to specify the first set of conditions now appears in the box.
7. Click the first line in the Select Operator column, and select **Contains**, if it is not already selected. Leave the first line in the Condition column blank.
8. Click  above the Step 2: String(s) to Exclude From the Results of Step 1 box.

   A line enabling you to specify the first set of conditions now appears in the box.
9. Click the first line in the Select Operator column, and select **Contains**.
10. In the first line of the Condition column, type the protein reference or sequence that you want to remove.
11. Click **Compile Database**.

    The compiling process creates the target FASTA file that excludes protein entries that match the condition.

FASTA Database Utilities Dialog Box Parameters

The following sections list the parameters on the pages of the FASTA Database Utilities dialog box.

- Add Protein References Page
- Compile FASTA Database Page
- Find Protein References Page

Add Protein References Page

Table 27 describes the parameters on the Add Protein References page of the FASTA Database Utilities dialog box.

Table 27. Add Protein References page parameters

Parameters	Description
FASTA File	Specifies the name of the FASTA file to modify.
Enter Description	Describes the protein sequence that you are appending to the selected FASTA file. The description should include the name of the protein and the source tissue.
Enter Protein Sequence	Specifies the amino acid sequence of the protein that you are appending to the selected FASTA file.
Add Entry	Appends and saves the protein sequence and reference to the currently selected FASTA file.

Compile FASTA Database Page

Table 28 describes the parameters on the Compile FASTA Database page of the FASTA Database Utilities dialog box.

Table 28. Compile FASTA Database page parameters (Sheet 1 of 3)

Parameters	Description
Original	Specifies the path and name of the source FASTA database that you are taking the information from. It must be an existing FASTA file.
Target	Specifies the path and name of the FASTA file that you are placing or adding the extracted information to.

Table 28. Compile FASTA Database page parameters (Sheet 2 of 3)

Parameters	Description
Target Database Options	Determines whether the extracted information is placed in a new FASTA file or is appended to an existing FASTA file. • Create/Replace (Default): Creates a new FASTA file where you will store the extracted information. • Append: Adds the extracted information to an existing FASTA file.
Ignore Case of Reference Strings	Disregards the case of the text in the Step 1: String(s) to Include box and the Step 2: String(s) to Exclude From the Results of Step 1 box.
Step 1: String(s) to Include:	Specifies the information to be extracted from the original file.
Active	Specifies the number of the condition set and allows you to select it.
Select Operator	Specifies the action to apply to the condition selected in the Condition column: • Starts With: Extracts information that begins with this string. • Does Not Start With: Extracts references that do not begin with this string. • Ends With: Extracts information that ends with this string. • Does Not End With: Extracts references that do not end with this string. • Contains: Extracts information that includes this string. • Does Not Contain: Extracts references that do not contain this string.
Condition	Specifies the condition that the information must meet in order to be extracted.
	Adds a line to the Step 1: String(s) to Include box.
	Deletes a line from the Step 1: String(s) to Include box.
Step 2: String(s) to Exclude From the Results of Step 1	Specifies the information to be excluded from the extracted results.
Active	Specifies the number of the condition set and allows you to select it.

Table 28. Compile FASTA Database page parameters (Sheet 3 of 3)

Parameters	Description
Select Operator	Specifies the action to apply to the condition selected in the Condition column: • Starts With: Excludes information that begins with this string. • Does Not Start With: Excludes information that does not begins with this string. • Ends With: Excludes information that ends with this string. • Does Not End With: Excludes information that does not end with this string. • Contains: Excludes information that includes this string. • Does Not Contain: Excludes information that does not include this string.
Condition	Specifies the condition that the information must meet in order to be excluded.
	Adds a line to the Step 2: String(s) to Exclude From the Results of Step 1 box.
	Deletes a line from the Step 2: String(s) to Exclude From the Results of Step 1 box.
Compile Database	Compiles and saves the changes to the FASTA file specified in the Target box.
Stop	Suspends the compilation.

Find Protein References Page

Table 29 describes the parameters on the Find Protein References page of the FASTA Database Utilities dialog box.

Table 29. Find Protein References page parameters (Sheet 1 of 2)

Parameters	Description
FASTA Database	Specifies the name of the FASTA file that you want to search for sequences or references in.
Search For	Specifies the amino acid sequence to search for.

Table 29. Find Protein References page parameters (Sheet 2 of 2)

Parameters	Description
Search In	Determines whether to search for the search string in the protein references or sequences. • References: Searches for the search string in the protein references. • Sequences: Searches for the specified amino acid sequence within the protein sequences.
Maximum Number of Matches Reported	Specifies the maximum number of matching references or sequences to report.
Start Search	Begins the search for the reference or sequence.
Save/Add Selected to Database	Adds a selected protein result row to a specified database.
Stop Search	Suspends the active search.
Reference: List of Operators	Specifies the operator for the condition: • Equals • Does Not Equal • Less Than • Less Than or Equal To • Greater Than • Greater Than or Equal To • Like • Matches Regular Expression • Starts With • Contains • Ends With • Does Not Start With • Does Not Contain • Does Not End With • Does Not Match • Not Like
Reference: List of Condition Settings	Custom: Opens the Enter Filter Criteria for Reference dialog box so that you can define additional conditions.
Clear Reference Filter icon ( )	Deletes the condition under “Reference” on the Find Protein References page.
Clear All Filters icon ( )	Deletes all conditions in the Enter Filter Criteria for Reference dialog box and the condition under “Reference” on the Find Protein References page.

Enter Filter Criteria for Reference Dialog Box Parameters

Table 30 describes the parameters on the Enter Filter Criteria for Reference dialog box.

Table 30. Enter filter criteria for Reference dialog box parameters

Parameters	Description
Operator	Specifies the operator for the condition. See "Find Protein References Page" on page 177 for a list of these operators.
Operand	Specifies the operand for the condition.
And conditions	Applies all the conditions that you have specified.
Or conditions	Applies one of the conditions that you have specified.
Add a condition	Adds a line to the Operator and Operand columns so you can add a new condition.
Delete condition	Removes a condition.

Managing FASTA Indexes

A FASTA index is a type of lookup table containing masses, theoretical peptide sequences, and associated proteins, which minimizes search time. The index lists all possible amino acid sequences that can be produced when an enzyme digests a protein or peptide. The peptide fragments are listed by molecular weight. The index stores information about every nominal mass, every peptide that has that mass, every protein that contains this peptide, and the location of its protein description in the FASTA file. Rather than read all protein sequences from the FASTA file, digest them in silico with the specified enzyme, calculate the mass of each peptide, and compare it to the given precursor mass, the Proteome Discoverer application looks for the specific mass in the FASTA index and uses it to find the peptides that have this mass and the associated proteins that contain the peptides.

For full enzymatic searches, the Proteome Discoverer application automatically creates FASTA indexes as they are needed. It does not automatically create FASTA indexes during semi-enzymatic or no-enzyme searches because these searches usually consume a large amount of space on a computer's hard disk. However, you can manually create FASTA indexes for these types of searches.

- Specifying the Location and Number of FASTA Indexes Stored
- Displaying the FASTA Indexes View
- Specifying the Columns to Display
- Automatically Creating a FASTA Index
- Manually Creating FASTA Indexes
- Controlling Automatic FASTA Index Removal

- Deleting a FASTA Index
- Changing Number and Location of Stored FASTA Indexes
- Removing FASTA Indexes When a FASTA File Is Deleted

Specifying the Location and Number of FASTA Indexes Stored

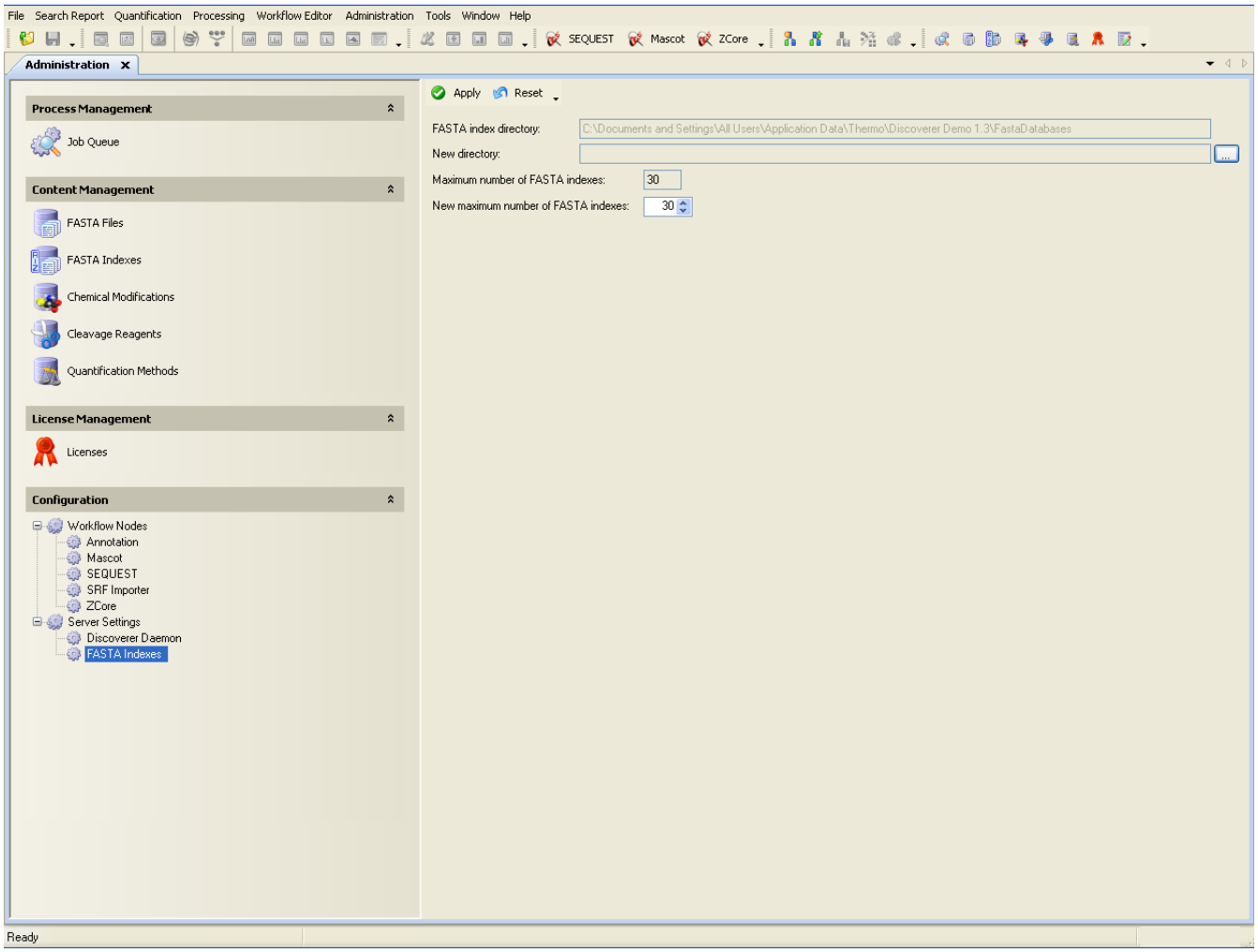
If you do not want to store the FASTA indexes in the default directory shown in Figure 114 on page 180, you can specify an alternate directory in the FASTA Indexes configuration view. You can also change the maximum number of FASTA indexes stored.

❖ To specify the location and number of the FASTA indexes stored

1. Choose **Administration > Server Settings > FASTA Indexes**.

   The configuration view shown in Figure 114 on page 180 appears.

Figure 114. FASTA Indexes configuration view

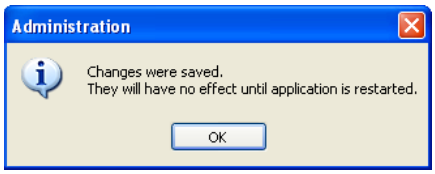


2. In the New Directory box, browse to the location of the folder to store the FASTA indexes in.
3. In the New Maximum Number of FASTA Indexes, box, select the maximum number of FASTA indexes to store.

   If you generate more FASTA indexes than the number to store in the New Maximum Number of FASTA Indexes box, the Proteome Discoverer application discards the difference from the oldest FASTA indexes the next time that you restart the application.
4. If you changed any settings, click Apply.

   The message box shown in Figure 115 appears:

   Figure 115. Administration message box

5. Click **OK**.

   Note Click Reset to return to the default values.
6. Restart your machine.

Displaying the FASTA Indexes View

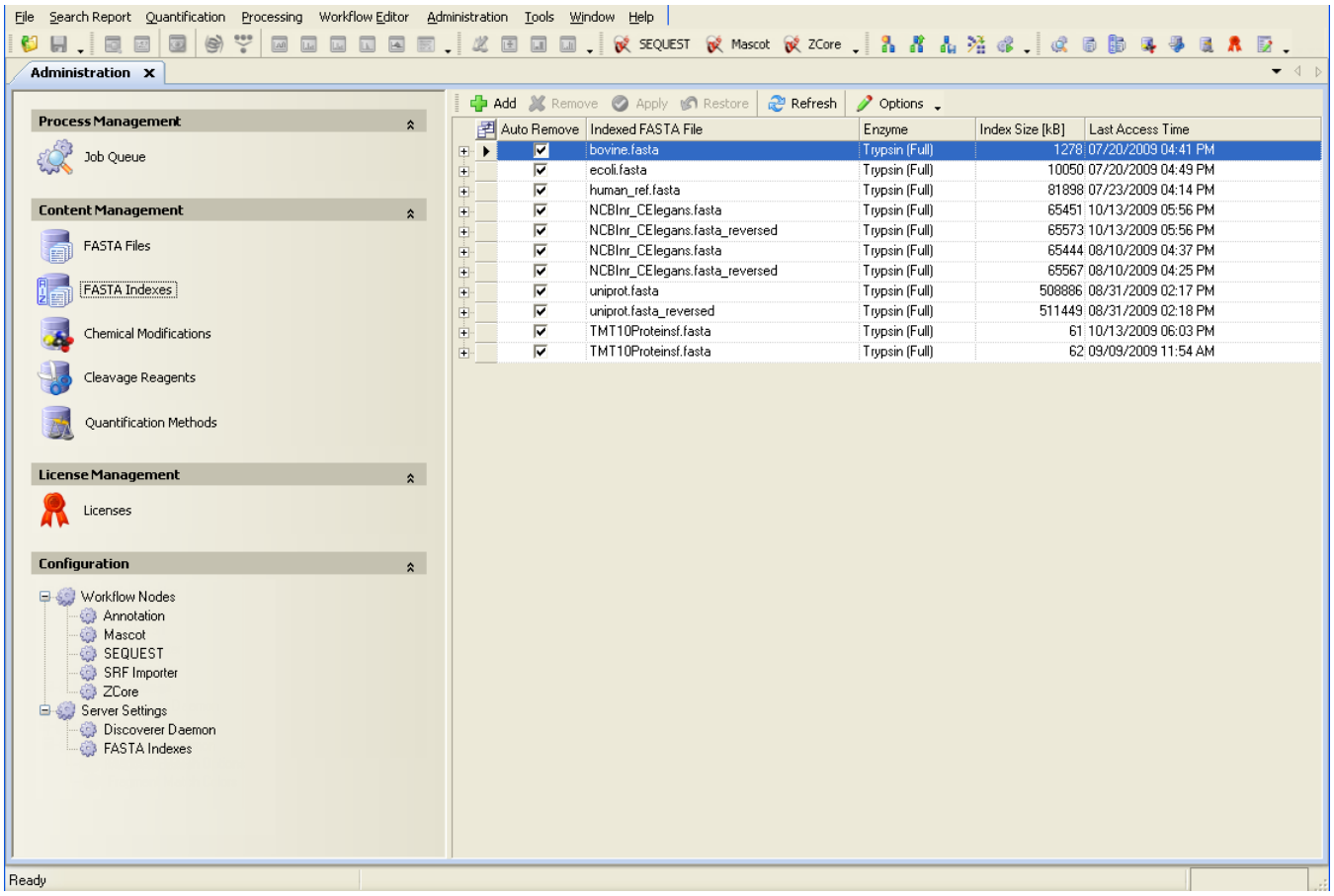
You can access FASTA indexes through the FASTA Indexes view.

❖ To display the FASTA Indexes view

1. Choose **Administration > Maintain FASTA Indexes** or click the **Maintain FASTA Indexes** icon, .

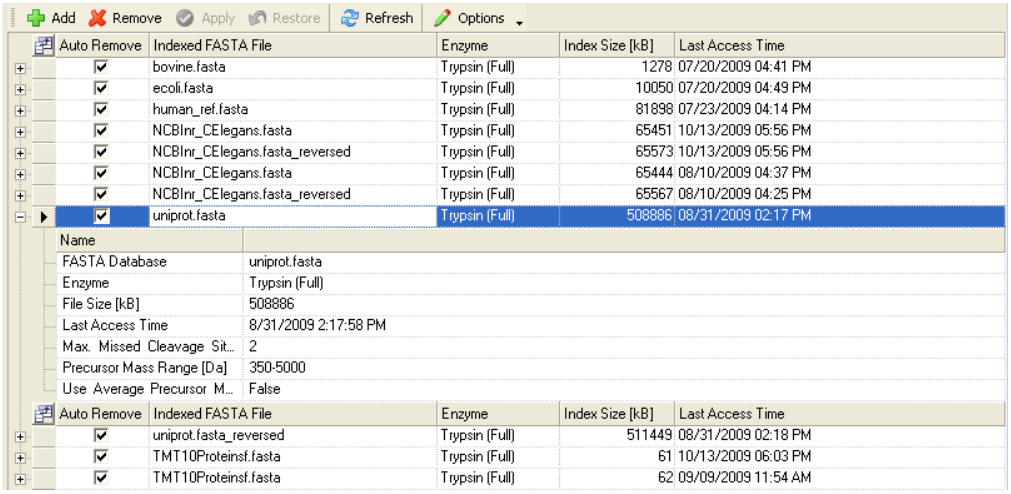
   The FASTA Indexes view appears, as shown in Figure 116.

Figure 116. FASTA Indexes view



2. Click the plus (+) sign to the left of a database name to vertically display the settings for that database, as shown for the uniprot.fasta database in Figure 117.

Figure 117. Database settings in the FASTA Indexes view



Specifying the Columns to Display

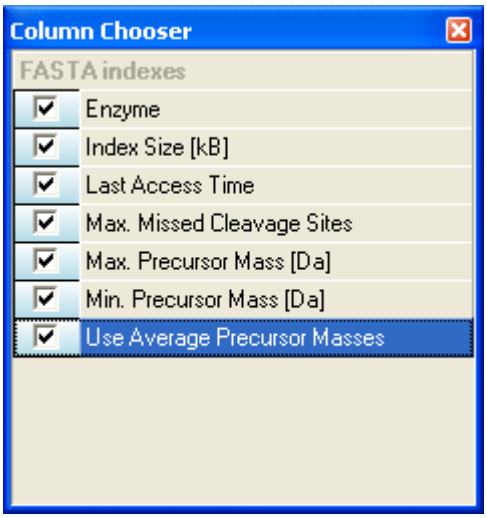
Use the Column Chooser to specify the columns that you want to display.

❖ To set the columns that you want to display

1. Click the **Column Chooser** icon, .
2. In the Column Chooser dialog box, shown in Figure 118, select the check boxes corresponding to the columns that you want to display in the FASTA Indexes view.

   The Proteome Discoverer application instantly makes the selected columns visible and the cleared columns invisible. For a description of these columns, see Table 31.

   Figure 118. Column Chooser dialog box in the FASTA Indexes view



FASTA Indexes View Parameters

The FASTA Indexes view on the Administration page contains the parameters shown in Table 31.

Table 31. FASTA Indexes view parameters (Sheet 1 of 3)

Parameters	Description
Add	Opens the FASTA Index Creator dialog box, shown in Figure 119 on page 186, so that you can add a FASTA index.
Remove	Removes the selected FASTA index and places it in the Deleted FASTA Indexes table in the FASTA Indexes view.

Table 31. FASTA Indexes view parameters (Sheet 2 of 3)

Parameters	Description
Apply	Applies the change that you made to the Auto Remove check box. If you clear the check box, the Proteome Discoverer application will no longer automatically remove the available FASTA indexes from memory after the number of indexes reaches the specified maximum.
Restore	Restores a deleted FASTA index and returns it to the FASTA Indexes table.
Refresh	Lists a new FASTA index in the FASTA Indexes view on the Administration page.
Options	Opens the FASTA Indexes Options dialog box, shown in Figure 123 on page 192, so you can specify the maximum number and location of the stored FASTA indexes.
Column Chooser	Opens the Column Chooser dialog box, shown in Figure 118 on page 183, so that you can select the columns to display in the FASTA Indexes view.
Auto Remove	Specifies whether the available FASTA indexes are removed from memory after the number of indexes reaches the specified maximum.
Indexed FASTA File	Displays the name of the FASTA database.
Enzyme	Specifies the name of the enzyme used to cleave the protein.
Index Size [kB]	Displays the size of the FASTA database index, in kilobytes.
Last Access Time	Displays the last time that the FASTA index was used. For example, it displays the last time that a SEQUEST search used the FASTA index.
Min. Precursor Mass [Da]	Specifies the minimum mass of the precursor ion to be processed, in daltons. The minimum value is 0.0 Da, and the maximum is 10 000.0 Da. The default is 350 Da.
Max. Precursor Mass [Da]	Specifies the maximum mass of the precursor ion to be processed, in daltons. The minimum value is 0.0 Da, and the maximum is 10 000.0 Da. The default is 5000 Da.

Table 31. FASTA Indexes view parameters (Sheet 3 of 3)

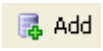
Parameters	Description
Max. Missed Cleavage Sites	Specifies the maximum number of internal cleavage sites per peptide fragment that is acceptable for the enzyme to miss when cleaving peptides during digestion. The minimum value is 0, and the maximum value is 12. The default is 2.
Use Average Precursor Masses	Determines whether the average mass is used to match the precursor ion: • True: Uses the average mass to match the precursor ion. • False: Uses the monoisotopic mass to match the precursor ion. The monoisotopic mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Automatically Creating a FASTA Index

The Proteome Discoverer application automatically creates FASTA indexes for a full enzymatic digestion during a SEQUEST or ZCore search, if an adequate FASTA index does not already exist. You can manually create a FASTA index for a semi-enzymatic or non-specific digestion (see "Manually Creating FASTA Indexes" on page 189).

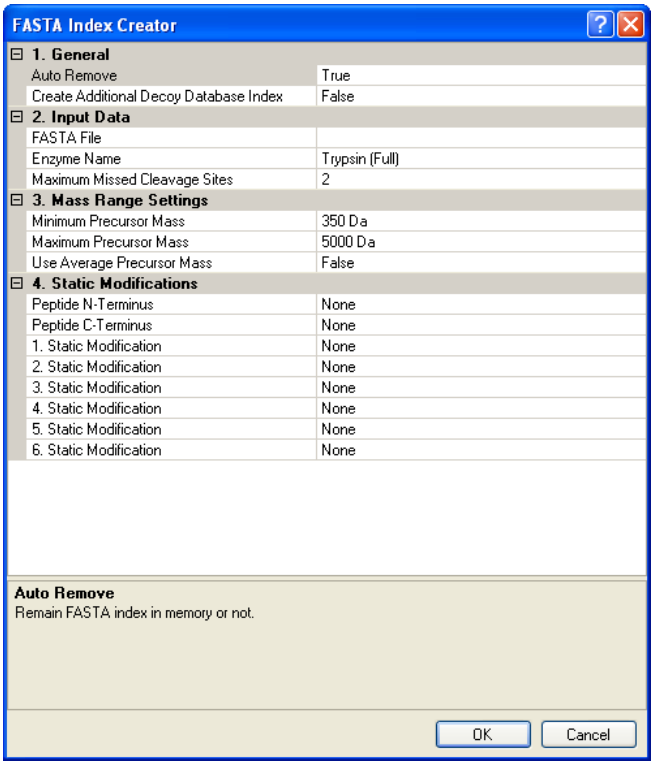
You can only create a specific FASTA index once.

❖ To automatically create a FASTA index

1. Choose **Administration > Maintain FASTA Indexes** or click the **Maintain FASTA Indexes** icon,  .
2. Click the **Add** icon,  .

   The FASTA Index Creator dialog box appears, as shown in Figure 119.

Figure 119. FASTA Index Creator dialog box



3. In the General section, specify whether the available FASTA indexes will be removed from memory after the number of indexes reaches the specified maximum.

   - True (Default): Automatically removes the FASTA indexes from memory.
   - False: Keeps the FASTA indexes in memory.

   For information about how the Proteome Discoverer application removes FASTA indexes after the maximum has been reached, see "Manually Creating FASTA Indexes" on page 189. For instructions on specifying the maximum number of indexes, see "Changing Number and Location of Stored FASTA Indexes" on page 191.

4. In the Input Data section, specify the basic information that the Proteome Discoverer application needs to create the index:

   - FASTA File: Select the FASTA database to be indexed from the list.
   - Enzyme Name: Select the enzyme used in the digestion from the list on the left (the enzymes on this list are set in the Cleavage Reagents window) and the type of digestion from the list on the right:
     - Full: Specifies a full enzymatic digestion.
     - Semi: Specifies semi-enzymatic digestion.
     - Unspecific: Specifies a non-specific digestion.
     - No Cleavages: Specifies that no cleavages occur.

- Maximum Missed Cleavage Sites: Specifies the maximum number of internal cleavage sites per peptide fragment that is acceptable for an enzyme to miss when cleaving peptides during digestion. Normally the digestion time is too short to enable the enzyme to cleave the peptide at all positions, so you must specify the number of missed positions in one resulting peptide fragment where the enzyme could cleave but did not.

  The minimum value is 0, and the maximum value is 12. The default is 2.

5. In the Mass Range Settings section, set the limits of the mass range of the singly charged precursor ion to be processed:

   - Minimum Precursor Mass: Specifies the minimum mass of the precursor ion. The minimum value is 0.0 Da, and the maximum value is 10000.0 Da. The default is 350 Da.
   - Maximum Precursor Mass: Specifies the maximum mass of the precursor ion. The minimum value is 0.0 Da, and the maximum value is 10000.0 Da. The default is 5000 Da.
   - Use Average Precursor Mass: Determines whether the average mass is used to match the precursor ion.
     - True: Uses the average mass to match the precursor ion.
     - False (Default): Uses the monoisotopic mass to match the precursor ion, which is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

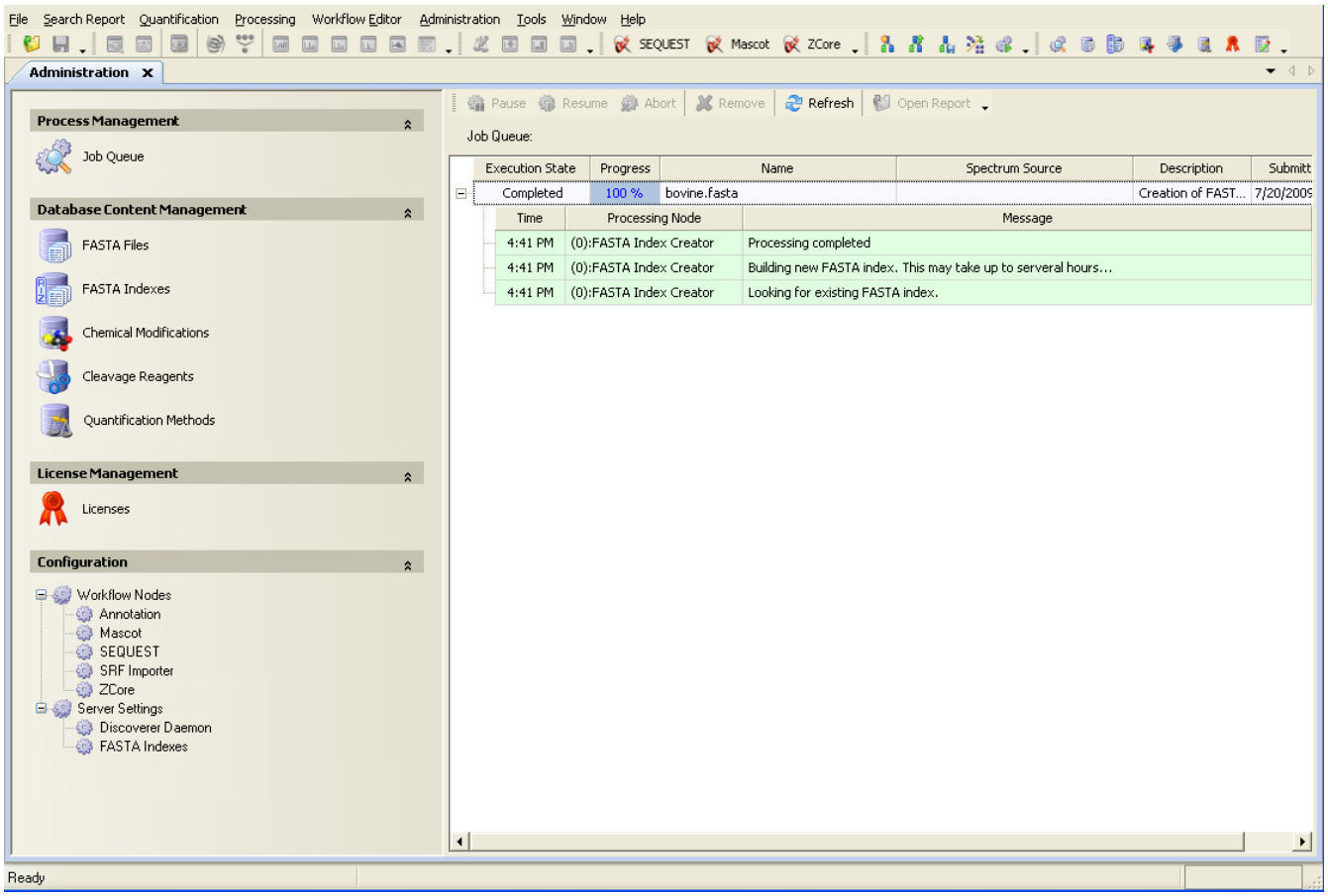
6. In the Static Modifications area, specify the static modifications that occur on the amino acid:

   - Peptide N-Terminus: Select the static modification that occurs on the N terminus of the peptide.
   - Peptide C-Terminus: Select the static modification that occurs on the C terminus of the peptide.
   - Static Modification: Select the static modification that occurs on the amino acid side chain.

7. Click **OK**.

   The Proteome Discoverer application starts creating the FASTA index, and the job queue appears, as shown in Figure 120.

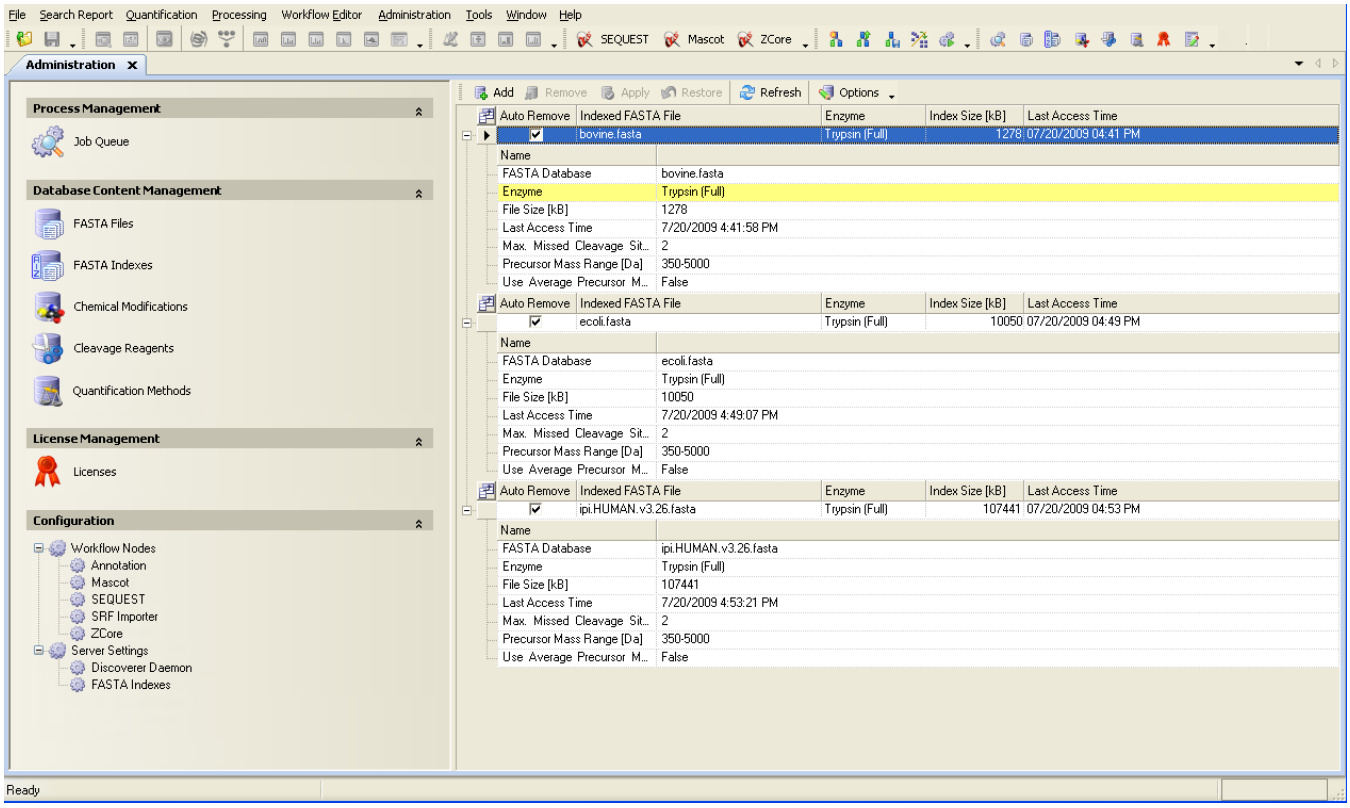
Figure 120. Creating a FASTA index



8. When the job finishes, choose **Administration > Maintain FASTA Indexes** or click the **Maintain FASTA Indexes** icon, , to display the FASTA Indexes view.
9. In the FASTA Indexes view, click the **Refresh** icon, Refresh.

   The new FASTA index appears in the FASTA Indexes view on the Administration page, as shown in Figure 121.

Figure 121. FASTA Indexes view



Manually Creating FASTA Indexes

As noted earlier, you can manually create FASTA indexes for semi-enzymatic or no-enzyme searches.

❖ To manually create a FASTA index

1. Follow the procedure in "Automatically Creating a FASTA Index" on page 185. Also, set the Create Additional Decoy Database Index parameter in the FASTA Index Creator dialog box to **True**, as shown in Figure 119 on page 186.

   The Proteome Discoverer application starts creating the FASTA index, and the job queue appears.

2. When the job finishes, choose **Administration > Maintain FASTA Indexes** or click the **Maintain FASTA Indexes** icon, , to display the FASTA Indexes view.

3. In the FASTA Indexes view, click the **Refresh** icon, Refresh.

   The new FASTA index appears in the FASTA Indexes view on the Administration page. The Proteome Discoverer application creates an index for the specified FASTA file and the decoy version of the FASTA file.

Controlling Automatic FASTA Index Removal

After the number of FASTA indexes reaches the specified maximum, the Proteome Discoverer application automatically removes from memory the number of FASTA indexes over the maximum. It first removes the oldest indexes (that is, the ones with the earliest access time). However, you can mark specific FASTA indexes so that they will not be removed from memory, even after the maximum is reached.

- To deactivate automatic FASTA index removal
- To activate automatic FASTA index removal

❖ To deactivate automatic FASTA index removal

1. In the FASTA Indexes view on the Administration page, clear the **Auto Remove** check box.

   The Apply icon now becomes available.

2. Click the **Apply** icon,  .

3. In the Remove FASTA indexes confirmation box, click **OK**.

❖ To activate automatic FASTA index removal

1. Select the **Auto Remove** check box.
2. Click the **Apply** icon,  .
3. In the Remove FASTA indexes confirmation box, click **OK**.

Deleting a FASTA Index

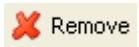
You can only delete FASTA indexes that have an Auto Remove check box selected.

- To delete a FASTA index
- To restore a deleted FASTA index

❖ To delete a FASTA index

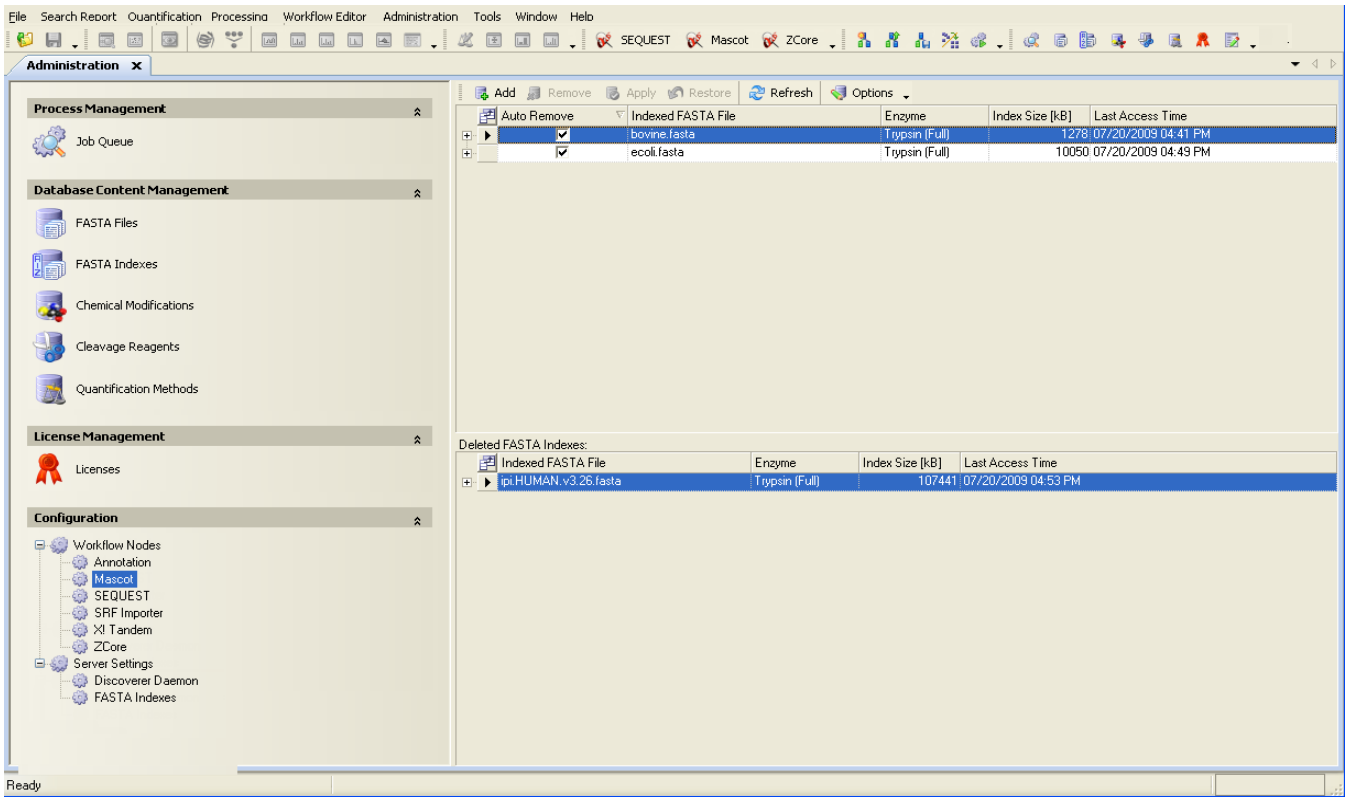
1. Be sure that the Auto Remove check box is selected for the index that you want to delete.
2. Select the index that you want to delete by clicking the box next to the plus (+) sign.

   The box now changes to the  arrow.

3. Click the  arrow.
4. Click the **Remove** icon,  .
5. Click **OK** in the Remove FASTA indexes confirmation box.

The name of the deleted index disappears from the FASTA Indexes table and reappears in a separate table called Deleted FASTA Indexes, as shown in Figure 122. It no longer appears in the FASTA Indexes table. However, because the FASTA index might be used in some calculations, its removal from the application only takes place the next time that the server starts.

Figure 122. Deleted FASTA Indexes table



❖ To restore a deleted FASTA index

1. In the Deleted FASTA Indexes table, select the deleted index by clicking the ▶ arrow.
2. Click the **Restore** icon, Restore.
3. In the Restore FASTA indexes confirmation box, click **OK**.

   The restored index appears in the FASTA Indexes table and disappears from the Deleted FASTA Indexes table.

Changing Number and Location of Stored FASTA Indexes

You can specify a new directory for storing the FASTA indexes and change the maximum number of FASTA indexes stored. The Proteome Discoverer application counts all FASTA indexes, even the indexes that cannot be automatically removed with the Auto Remove option.

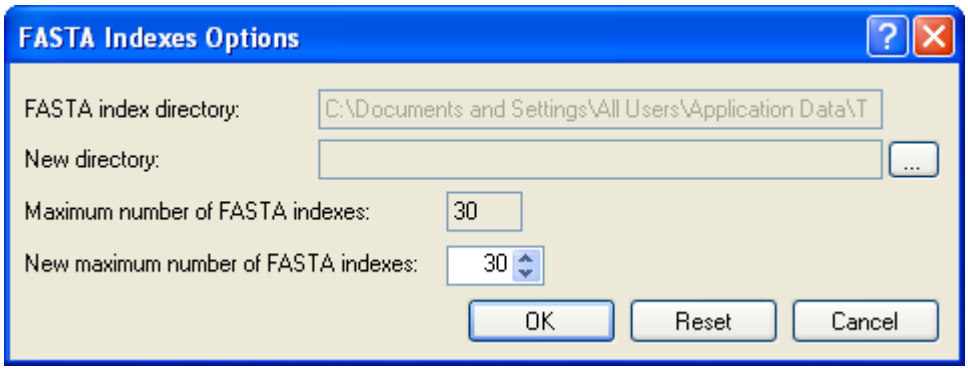
- To change the number and location of stored FASTA indexes
- To reset the changes made in a previous FASTA index session

❖ To change the number and location of stored FASTA indexes

1. Click the **Options** icon, Options .

   The FASTA Indexes Options dialog box appears, as shown in Figure 123.

   Figure 123. FASTA Indexes Options dialog box



   Note Another way to access these options is to choose Administration > Configuration and click FASTA Indexes in the Server Settings area.

   The FASTA Indexes Options dialog box contains two read-only parameters:

   - The FASTA Index Directory box displays the name of the current directory where the FASTA indexes are saved.
   - The Maximum Number of FASTA Indexes box displays the current maximum number of FASTA indexes allowed.

2. In the New Directory box, browse to the directory where you want to store the FASTA indexes.

   You can change the directory only if the server runs on the local machine.

3. In the New Maximum Number of FASTA Indexes box, type the new maximum number of FASTA indexes allowed.

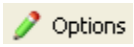
4. Click **OK**.

5. In the FASTA index settings confirmation box, click **OK**.

After you confirm the changes, the Proteome Discoverer application saves them, but the changes are only executed the next time that the server starts. You can undo the changes made since the last time that the server started and before the next time that the server starts, even though you clicked OK in the FASTA Indexes Options dialog box and closed it. For example, when you change the location of the directory in the FASTA Indexes Options dialog box,

click OK, and close the dialog box, the server moves all FASTA indexes to the new target directory when the server restarts. But if you reinvoke the dialog box and click Reset before restarting the server, the changes that you made previously are deleted, and the directory reverts to its previous location.

❖ **To reset the changes made in a previous FASTA index session**

1. Click the **Options** icon, 

.

   The FASTA Indexes Options dialog box appears, as shown in Figure 123.

2. Click **Reset**.

Removing FASTA Indexes When a FASTA File Is Deleted

When you or the Proteome Discoverer application deletes a FASTA file, the application removes the FASTA indexes belonging to the deleted FASTA file the next time that the server starts.

FASTA Index Creator Dialog Box Parameters

Table 32 describes the parameters on the FASTA Index Creator dialog box.

Table 32. FASTA Index Creator dialog box parameters (Sheet 1 of 3)

Parameters	Description
General	Sets general options for the creation of the FASTA index.
Auto Remove	Specifies whether the available FASTA indexes will be removed from memory after the number of indexes reaches the specified maximum. • (Default) True: Automatically removes the FASTA indexes from memory. • False: Keeps the FASTA indexes in memory.
Create Additional Decoy Database Index	Determines whether the Proteome Discoverer application creates an index for the specified FASTA file and an index for the decoy version of the FASTA file. • True: Creates an index for the specified FASTA file and an index for the decoy version of the FASTA file. • (Default) False: Does not create an index for the specified FASTA file nor an index for the decoy version of the FASTA file.

Table 32. FASTA Index Creator dialog box parameters (Sheet 2 of 3)

Parameters	Description
Input Data	Specifies the basic information that the Proteome Discoverer application needs to create the index.
FASTA File	Specifies the name of the FASTA database.
Enzyme Name	Specifies the name of the enzyme used for digestion and the type of digestion. The list on the right specifies the type of digestion, which can be one of the following: • Full: Specifies a full enzymatic digestion. • Semi: Specifies semi-enzymatic digestion. • Unspecific: Specifies a non-specific digestion. • No Cleavages: Specifies that no cleavages occur.
Maximum Missed Cleavage Sites	Specifies the maximum number of internal cleavage sites per peptide fragment that is acceptable for an enzyme to miss when cleaving peptides during digestion. The minimum value is 0, and the maximum value is 12. The default is 2.
Mass Range Settings	Sets the limits of the mass range of the singly charged precursor ion to be processed.
Minimum Precursor Mass	Specifies the minimum mass of the precursor ion. The minimum value is 0.0 Da, and the maximum value is 10000.0 Da. The default is 350 Da.
Maximum Precursor Mass	Specifies the maximum mass of the precursor ion. The minimum value is 0.0 Da, and the maximum value is 10000.0 Da. The default is 5000 Da.
Use Average Precursor Mass	Determines whether the average mass is used to match the precursor ion. • True: Uses the average mass to match the precursor ion. • False (Default): Uses the monoisotopic mass to match the precursor ion, which is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Static Modifications	Specifies the static modifications that occur on the amino acid.
Peptide N-Terminus	Specifies the static modification that occurs on the N terminus of the peptide.

Table 32. FASTA Index Creator dialog box parameters (Sheet 3 of 3)

Parameters	Description
Peptide C-Terminus	Specifies the static modification that occurs on the C terminus of the peptide.
Static Modification boxes	Specifies the static modification that occurs anywhere on the amino acid side chain.

FASTA Indexes Options Dialog Box Parameters

Table 33 describes the parameters on the FASTA Indexes Options dialog box.

Table 33. FASTA Indexes Options dialog box parameters

Parameters	Description
FASTA Index Directory	Displays the path and name of the current directory where the FASTA indexes are stored. This parameter is read-only.
New Directory	Specifies the new path and name of the directory to store the FASTA indexes in.
Maximum Number of FASTA Indexes	Displays the current maximum number of FASTA indexes allowed. This parameter is read-only.
New Maximum Number of FASTA Indexes	Specifies the new maximum number of FASTA indexes allowed.
Reset	Resets the changes that you made in the FASTA Indexes Options dialog box to their values the last time the server started and prevents the server from implementing the changes when it starts next.

Chemical Modifications

You can update the chemical modifications that you use to conduct a peptide identification search. The available modifications are defined in the Chemical Modifications view on the Administration page that is opened by choosing Administration > Maintain Chemical Modifications. Use this view to customize the chemical modifications that you use to do your search. You can import a new list or the latest UNIMOD list. You can also modify the chemical modification list by adding amino acids to the modifications, creating new modifications, or activating or deactivating existing modifications.

Note A modification must be active to be usable during a search.

The Proteome Discoverer application offers two types of modifications, dynamic and static.

Dynamic Modifications

Dynamic modifications, also known as variable amino acid modifications, are modifications that might or might not be present. They are mainly used for determining post-translational modifications (PTMs). For example, some phosphorylated peptide serines are modified, and some are not modified.

You can set the parameters for a dynamic search on the Modifications page of the SEQUEST, Mascot, and ZCore search wizards. For instructions on setting these parameters in these wizards, see Figure 43 on page 79 and the steps that follow it.

Static Modifications

Static modifications apply the same specific mass to all occurrences of that named amino acid, as in an exhaustive chemical modification.

A static modification might result from derivatization or isotopic labeling of an amino acid. For example, a carboxymethylated cysteine has a delta mass of 58.005479, which is added to each cysteine residue appearing in a protein.

In static searches, the Proteome Discoverer application assumes that every amino acid residue will be modified in that way. Constant mass is changed. The search wizards perform static modification searches by adding the specified constant value to the mass of the specified amino acid.

You can set the parameters for a static search on the Modifications page of the SEQUEST, Mascot, and ZCore search wizards. For instructions on setting these parameters in the search wizards, see Figure 43 on page 79 and the steps that follow it.

Opening the Chemical Modifications View

The Chemical Modifications view is an advanced feature of the Proteome Discoverer application. You use it to build and maintain the static and dynamic modifications data that is available when you define your search settings.

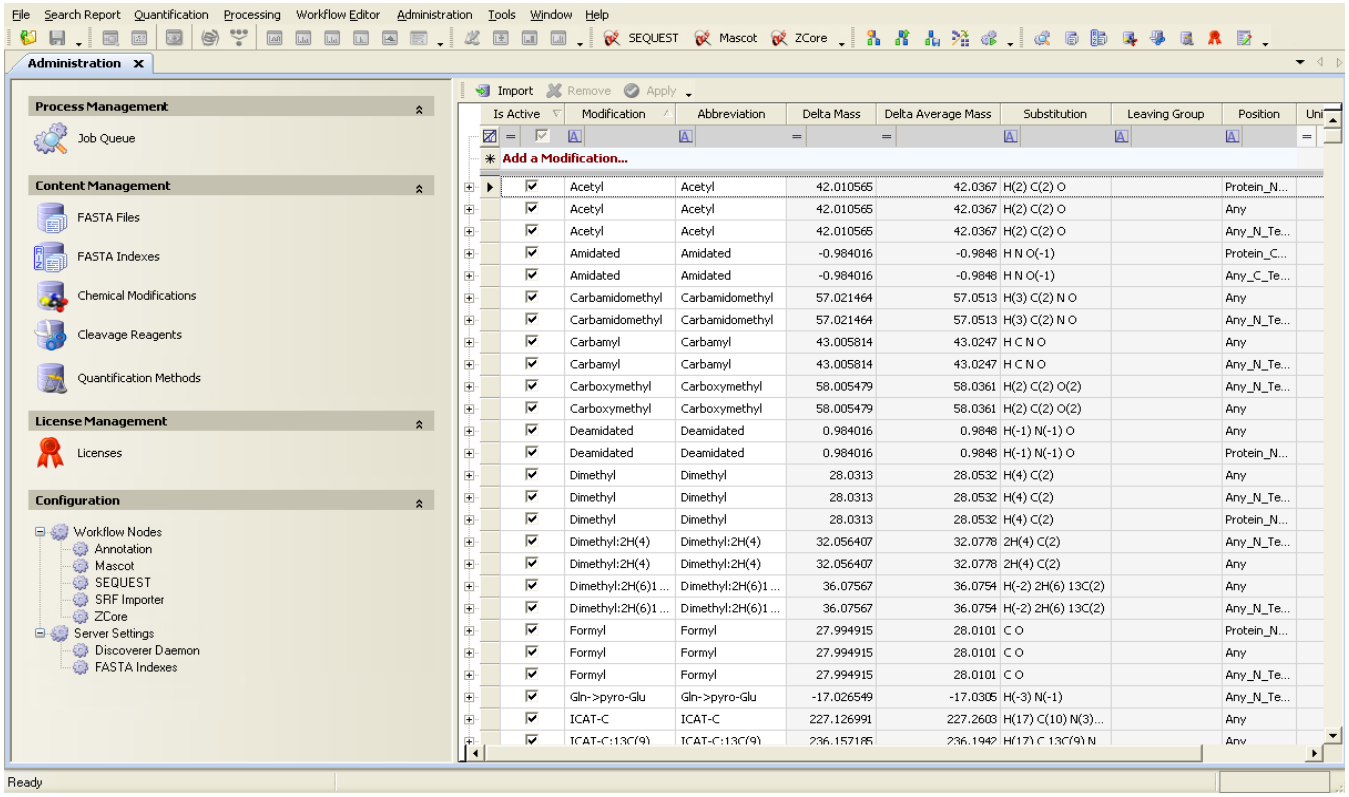
In the Chemical Modifications view, you can explore the default types of modifications and their corresponding amino acids. It contains the modification's delta mass, amino acids, and substitutions. By using the Chemical Modifications view, you can add amino acids to existing modifications and create new modifications.

❖ To open the Chemical Modifications view

1. Choose **Administration > Maintain Chemical Modifications**, or click the **Maintain Chemical Modifications** icon,  , either on the toolbar or on the Administration page.

   The Chemical Modifications view appears on the Administration page, as shown in Figure 124. The amino acids listed are those where the modifications can appear.

Figure 124. Chemical Modifications view



2. Click + to the left of each modification row to see the amino acids that the modification is found on, the letter abbreviation of this amino acid, and the modification type or category. Figure 125 shows an example of the information given for the Acetyl modification. Table 34 lists the available modification categories.

Figure 125. Displaying modification information for Acetyl

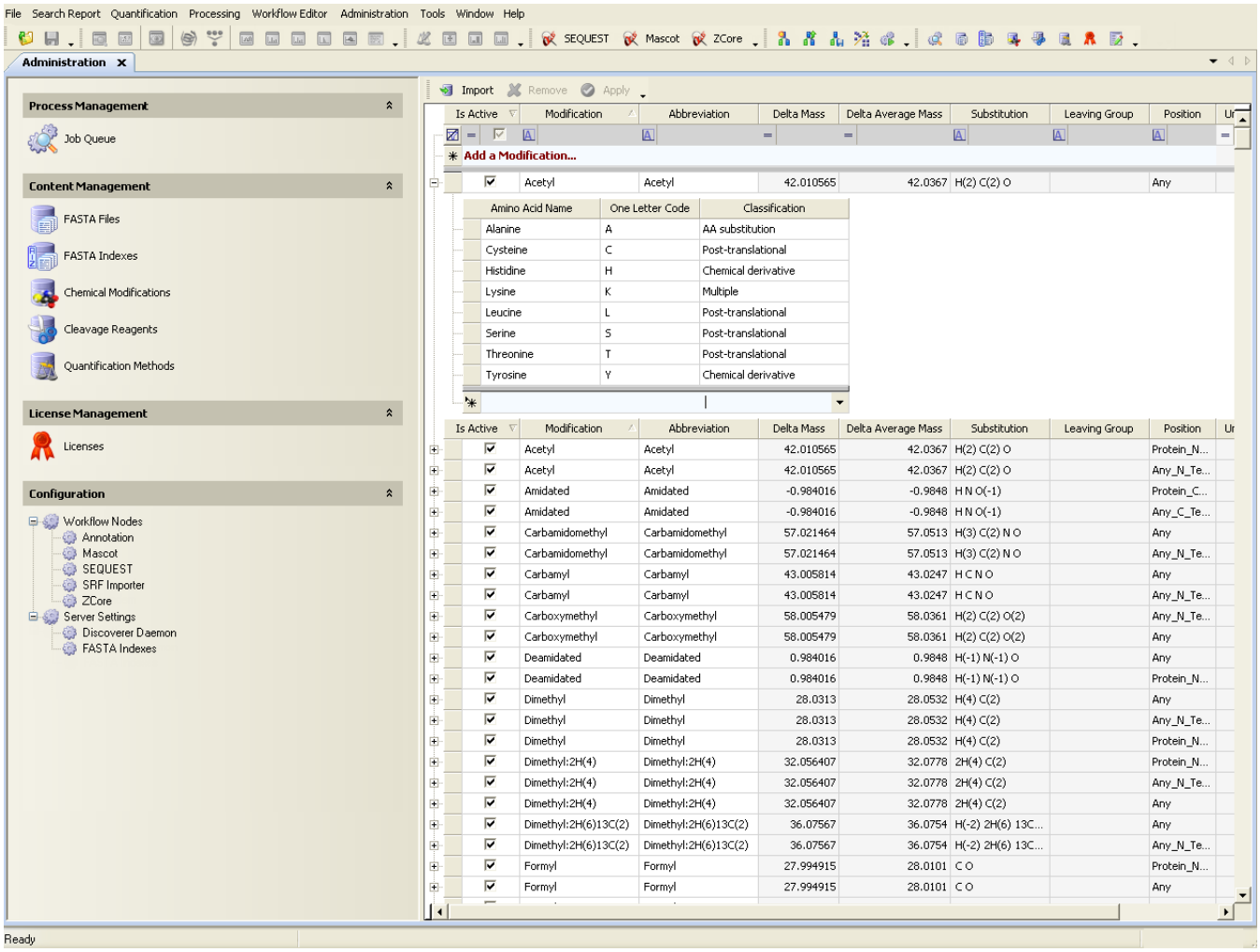


Table 34. Available modification categories (Sheet 1 of 2)

Classification	Description
Post-translational	Protein modification after translation (*in vivo*)
Co-translational	Amino acid modified in translation (for example, myristyl glycine)
Pre-translational	Amino acid modified before integration into a protein (for example, formyl methionine)
Chemical derivative	Chemically induced modification (for example, during sample preparation)
Artifact	Modification made during sample preparation
N-linked glycosylation	Glycosylation (in vivo)
O-linked glycosylation	Glycosylation (in vivo)

Table 34. Available modification categories (Sheet 2 of 2)

Classification	Description
Other glycosylation	Glycosylation (in vivo)
Synthetic peptide protection group	Protection group used in chemical peptide synthesis (for example, trityl (triphenylmethyl))
Isotopic label	Label for quantification
Non-standard residue	Amino acid derivative like selenomethionine
Multiple	More than one classification possible
AA substitution	Amino acid replaced by another amino acid (mutation)
Other	Modification not fitting into another category

The Proteome Discoverer application automatically imports the classifications from unimod.org, the protein modifications online database for mass spectrometry applications. You can also manually define your own classifications.

Chemical Modifications View Parameters

Table 35 describes the parameters on the Chemical Modifications view.

Table 35. Chemical Modifications view parameters (Sheet 1 of 2)

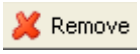
Parameter	Description
	Imports chemical modifications from a local file or updated versions from unimod.org.
	Deletes the selected amino acid or chemical modification.
	Saves a new modification or updates changes to an existing modification.
Is Active	Makes modifications available for search parameters.
Modification	Displays the type of modification.
Abbreviation	Displays the abbreviation of the modification.
Delta Mass	Displays the monoisotopic mass difference associated with the modification.
Delta Average Mass	Displays the average mass difference associated with the modification.
Substitution	Displays the chemical group substitution that occurs with the modification.
Leaving Group	Displays the chemical group that is lost.

Table 35. Chemical Modifications view parameters (Sheet 2 of 2)

Parameter	Description
Position	Displays the location where the modification might occur. The selections are Any, Any_C_Terminus, Any_N_Terminus, Protein_C_Terminus, and Protein_N_Terminus.
Unimod Accession No.	Refers to the accession number assigned to the modification by unimod.org. Modifications imported from unimod.org have an accession number greater than 0. User-added modifications have an accession number of 0.

Adding Chemical Modifications

You can create new chemical modifications and add them to the Chemical Modifications view. For example, you might have a new or experimental label that you want to add to the list of chemical modifications.

- To add a new chemical modification
- To update an existing chemical modification

❖ To add a new chemical modification

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears, as shown in Figure 124 on page 197.

2. Click the **Add a Modification** heading.

   An empty row appears, as shown in Figure 126.

Figure 126. Adding a row in the Chemical Modifications view

Is Active	Modification	Abbreviation	Delta Mass	Delta Average Mass	Substitution	Leaving Group	Position	Unimod Accession No.
☑			1	1			Any	0
☑	Acetyl	Acetyl	42.010565	42.0367	H(2) C(2) O		Any	1
☑	Acetyl	Acetyl	42.010565	42.0367	H(2) C(2) O		Protein_N...	1
☑	Acetyl	Acetyl	42.010565	42.0367	H(2) C(2) O		Any_N_Te...	1
☑	Amidated	Amidated	-0.984016	-0.9848	H N O(-1)		Protein_C...	2
☑	Amidated	Amidated	-0.984016	-0.9848	H N O(-1)		Any_C_Te...	2
☑	Carbamidomethyl	Carbamidomethyl	57.021464	57.0513	H(3) C(2) N O		Any	4
☑	Carbamidomethyl	Carbamidomethyl	57.021464	57.0513	H(3) C(2) N O		Any_N_Te...	4
☑	Carbamyl	Carbamyl	43.005814	43.0247	H C N O		Any	5
☑	Carbamyl	Carbamyl	43.005814	43.0247	H C N O		Any_N_Te...	5

3. In the empty row, enter the name of the modification, the delta masses, the chemical substitution, the chemical group that is leaving, the position, and the abbreviations of the modifications.

4. If you select Any in the Position column, a message box opens to inform you that you must specify which amino acids (target amino acids) will possibly have the modification. For instructions on this procedure, see "Adding Amino Acids" on page 201.
5. To accept the new modifications, click the **Apply** icon, 

.
6. Add an amino acid to the modifications. See "Adding Amino Acids" on page 201.

❖ To update an existing chemical modification

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears, as shown in Figure 124 on page 197.
2. In the Modification column, click the cell that you want to update**.**
3. Type your changes for the delta masses, the substitution, the group that it is leaving, the position, or the abbreviations of the modifications.

   For chemical modifications that you add yourself, you can edit any column except the Unimod Accession No. column. The Unimod Accession No. column identifies these modifications by a zero. For chemical modifications that you import from UNIMOD, you can edit only the Modification and Abbreviation columns. UNIMOD chemical modifications are identified by a number greater than zero in the Unimod Accession No. column.

   Columns that you can edit activate an edit button when you click them. Columns that you cannot edit display a gray background.
4. To accept the changes, click the **Apply** icon, 

.

Adding Amino Acids

You can add amino acids to a modification that has been set up for any position.

❖ To add an amino acid to a modification

1. Choose **Administration > Maintain Chemical Modifications**.

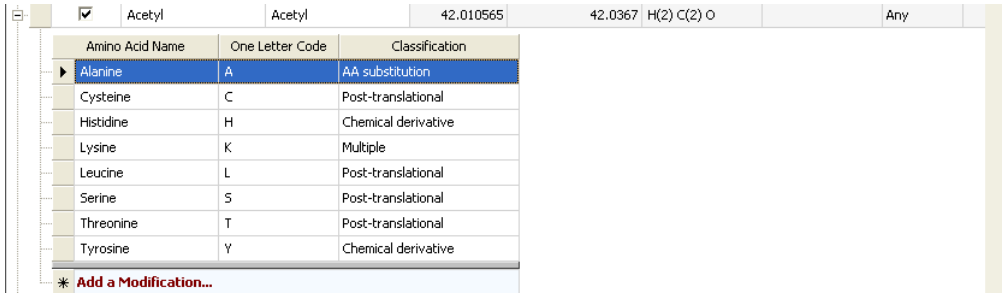
   The Chemical Modifications view appears, as shown in Figure 124 on page 197.
2. Click **+** to the left of the modification row that you want to update.

   The row must display Any in the Position column.

   The list of classifications now appears, as shown in Figure 125 on page 198.
3. Click the **Add a Modification** line below the list of amino acids.

   Figure 127 shows this line.

Figure 127. Adding an amino acid to a modification



An empty row appears.

4. In the empty row, select the amino acid from the list in the Amino Acid Name column.

   The amino acid and the one letter abbreviation appear.

5. From the list in the Classification column, select the type of modification.

6. To save the modifications, click the **Apply** icon, Apply.

When you reimport data from unimod.org, the Proteome Discoverer application retains the modification that you added. However, if you want to change the classification of an amino acid, you must do so before reimporting the Unimod data. After you import the Unimod data, the only way to change the classification is to delete the amino acid and re-add it with another classification.

Deleting Chemical Modifications

You can remove chemical modifications from the Chemical Modifications view.

❖ To delete a modification

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears, as shown in Figure 124 on page 197.

2. Select the row of the modification that you want to delete.

3. Click the **Remove** icon, Remove.

4. In the Delete Row dialog box, click **Yes**.

   The row is removed from the chemical modifications table.

Importing Chemical Modifications

You can import chemical modifications from a local file or obtain an updated version from unimod.org, a public domain database.

When you install the Proteome Discoverer application, it automatically imports accessions from unimod.org as chemical modifications.

- To import chemical modifications from a local file
- To import chemical modifications from unimod.org

❖ To import chemical modifications from a local file

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears, as shown in Figure 124 on page 197.

2. Click the **Import** icon, Import .

3. In the Import From list of the Import Modifications dialog box, select **Local File**.

4. In the adjacent box, click … to browse for your file, or type the name and path of the file in the box.

5. To overwrite an existing upload, select the **Overwrite Existing** check box.

6. Click **Import**.

   A status message appears.

7. When the upload is complete, click **Close**.

❖ To import chemical modifications from unimod.org

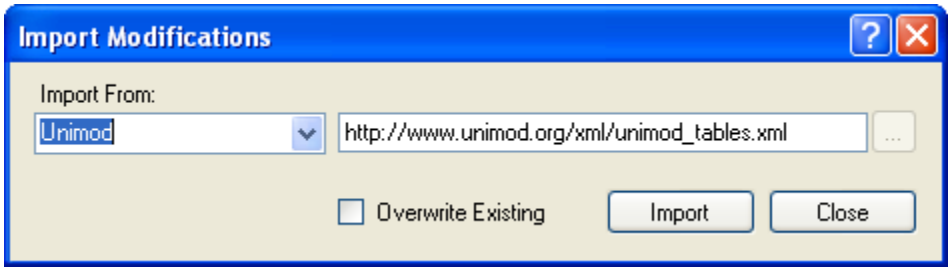
1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears, as shown in Figure 124 on page 197.

2. Click the **Import** icon, Import .

   The Import Modifications dialog box appears, as shown in Figure 128.

   Figure 128. Import Modifications dialog box



3. In the Import From list, select **Unimod**.

   The UNIMOD URL appears in the adjacent box.

4. To overwrite an existing upload, select the **Overwrite Existing** check box.

5. Click **Import**.

   A status message appears.

6. When the upload is complete, click **Close**.

For chemical modifications imported from unimod.org, you can only edit the Is Active, Modification, and Abbreviation columns. You do not have access to the Delta Mass, Delta Average Mass, Substitution, Leaving Group, Position, and UNIMOD Accession No. columns. Chemical modifications imported from unimod.org have a number greater than zero in the Unimod Access No. column.

If you select the Overwrite Existing check box, the Proteome Discoverer application does the following when it imports chemical modifications from unimod.org:

- Updates the columns that are inaccessible to you.
- Updates the names and the abbreviations of the modifications.
- Adds any new amino acids found in unimod.org.
- Adds any amino acids that you removed if they are defined in unimod.org.
- Removes any amino acids that you added if they are defined in unimod.org.

If you do not select the Overwrite Existing check box, the Proteome Discoverer application performs the same tasks as it does during installation:

- Updates the columns that are inaccessible to you.
- Leaves the modification name and abbreviation unchanged.
- Adds any new amino acids found in unimod.org.
- Adds any amino acids that you removed if they are defined in unimod.org.
- Leaves unchanged any amino acids that you added.

Deleting Amino Acids

You can also delete amino acids from chemical modifications.

❖ To delete an amino acid from a chemical modification

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears, as shown in Figure 124 on page 197.

2. Click **+** to the left of the modification row that you want to delete.

   The row expands and the associated amino acids appear.

3. Select the amino acid row that you want to delete.

4. Click the **Remove** icon,  .
5. In the Delete Row dialog box, click **Yes**.

   The row is removed from the chemical modifications table.

Using the Qual Browser Application

The Proteome Discoverer application includes the Qual Browser application, which you can use to examine spectra and chromatograms in detail. With the Qual Browser application, you can view the entire ion chromatogram and browse individual precursor and MS^n data. You can filter the results in a variety of ways, for example, to produce a selected ion chromatogram. The Qual Browser application automatically displays the elemental composition, theoretical mass, delta values, and ring and double-bond (RDB) equivalents for your high-resolution data. (RDB equivalents measure the number of unsaturated bonds in a compound and limit the calculated formulas to only those that make sense chemically.)

You must have the Xcalibur data system installed to use the Qual Browser application. For information about using the Qual Browser application, refer to the *Thermo Xcalibur Qualitative Analysis User Guide*.

You must also have a search results file open and a specific peptide or search input row selected before the Qual Browser application becomes available. If you are viewing the Administration page, the Qual Browser application does not open a raw file.

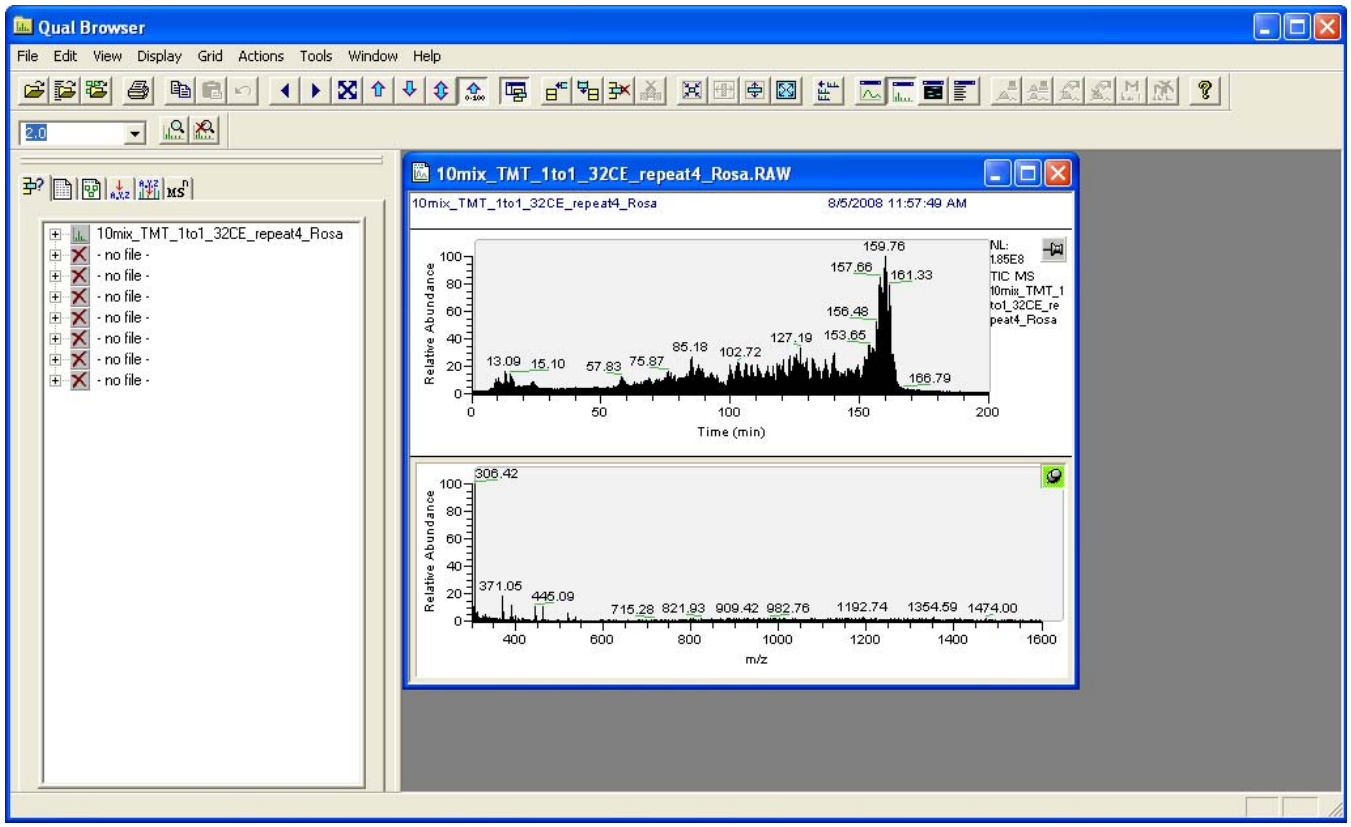
❖ To open the Qual Browser application

1. Choose **Tools > Open QualBrowser**, click the **Qual Browser** icon,  , or press CTRL+SHIFT+B to open the Spectrum window.

   Note You must have a search results (MSF) file open and selected before the Open QualBrowser command becomes available on the Tools menu. In addition, the Open QualBrowser command is available only when you select a single peptide or search input item first. You cannot use QualBrowser if the original raw file or files are missing.

   The Qual Browser application opens, as shown in Figure 129.

Figure 129. The Qual Browser application window



2. Right-click and choose **Display Options** from the shortcut menu.
3. To automatically annotate your peaks with the elemental composition, theoretical mass, RDB equivalent, or mass delta, click the **Composition** tab and select the labels for display.

Customizing Cleavage Reagents

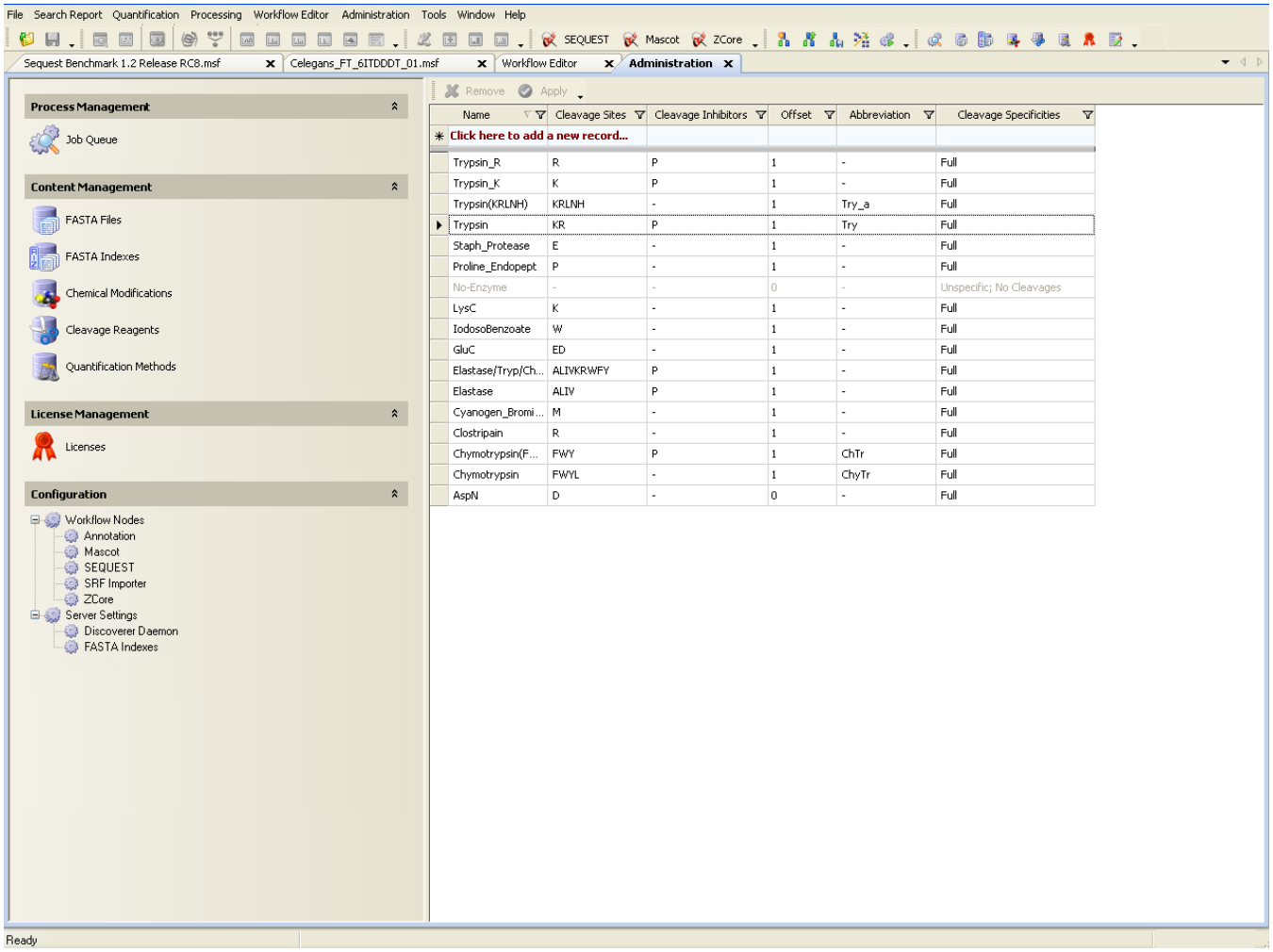
In the Cleavage Reagents view, you can explore the default types of reagents and their corresponding settings. You can also add, remove, and modify the reagents and their corresponding settings. The Cleavage Reagents view contains the cleavage sites, cleavage inhibitors, abbreviations, and cleavage specificities.

❖ To display the Cleavage Reagents view

- Choose **Administration > Maintain Cleavage Reagents**, or click the **Maintain Cleavage Reagents** icon, , on the toolbar or on the Administration page.

  The Cleavage Reagents view appears, as shown in Figure 130.

Figure 130. Cleavage Reagents view



Cleavage Reagents View Parameters

Table 36 describes the parameters on the Cleavage Reagents view.

Table 36. Cleavage Reagents view parameters (Sheet 1 of 4)

Parameter	Description
Remove	Deletes the highlighted cleavage reagent.
Apply	Saves any changes made to the cleavage reagents.

Table 36. Cleavage Reagents view parameters (Sheet 2 of 4)

Parameter	Description
Name	Specifies the name of the reagent used for the protein digestion. • All: Returns the filtered search results to the results that were first loaded. • Custom: Opens the Custom Filter dialog box, shown in Figure 131 on page 212. • Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked. • List of enzymes: Lists the specific names of all available enzymes that act as reagents.
Cleavage Sites	Specifies the position (amino acid) at which to cleave the sequence. • All: Returns the filtered search results to the results that were first loaded. • Custom: Opens the Custom Filter dialog box, shown in Figure 131 on page 212. • Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked. • List of amino acids: Lists the names of amino acids at which the enzyme cleaves the peptide.

Table 36. Cleavage Reagents view parameters (Sheet 3 of 4)

Parameter	Description
Cleavage Inhibitors	Specifies the amino acids that block cleavage when adjacent to the cleavage site. • All: Returns the filtered search results to the results that were first loaded. • Custom: Opens the Custom Filter dialog box, shown in Figure 131 on page 212. • Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked. • -: No amino acids will inhibit cleavage. • P: Adjacent proline amino acids inhibit cleavage
Offset	Specifies at which side of the given amino acids the protease should cleave: • All: Returns the filtered search results to the results that were first loaded. • Custom: Opens the Custom Filter dialog box, shown in Figure 131 on page 212. • Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked. • 0: The protease cleaves at the N-terminal side of the specified cleavage sites. • 1: The protease cleaves at the C-terminal side of the specified cleavage sites.

Table 36. Cleavage Reagents view parameters (Sheet 4 of 4)

Parameter	Description
Abbreviation	Specifies the user-defined abbreviation. • All: Returns the filtered search results to the results that were first loaded. • Custom: Opens the Custom Filter dialog box, shown in Figure 131 on page 212. • Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked. • List of abbreviated enzymes: Lists the existing user-defined abbreviations.
Cleavage Specificities	Selects the cleavage specificity. You can select more than one cleavage specificity. The available cleavage specificities are as follows: • All: Returns the filtered search results to the results that were first loaded. • Custom: Opens the Custom Filter dialog box, shown in Figure 131 on page 212. • Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked. • Full: Every cleavage is at the specified cleavage site. • Semi: Only one end of the sequence must have the specified cleavage. • Semi (N-Term): Only the N-terminal side of the sequence must have the specified cleavage. • Semi (C-Term): Only the C-terminal side of the sequence must have the specified cleavage. • Unspecific: Non-specific cleavage site is specified. • No Cleavages: No cleavage site is specified.
Click here to add a new record	Opens a row in the reagents table so you can enter the name and information about a new cleavage reagent.

Adding a Cleavage Reagent

❖ **To add a new cleavage reagent**

1. Click the Name column cell and click **Click Here To Add a New Record**.
2. Modify the default values in the row of that new reagent.
3. Click **Apply**.

Deleting a Cleavage Reagent

❖ **To delete a cleavage reagent**

1. Click the box in the * column next to the row that you want to delete.
2. Click **Delete**.
3. Click **Yes** in the confirmation box that appears.

Modifying a Cleavage Reagent

❖ **To modify a cleavage reagent**

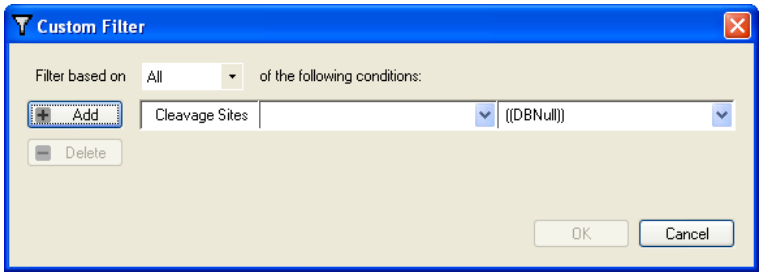
1. Click in the column for the reagent you want to modify, select the current contents, and enter the new information.
2. Click **Apply**.

Filtering Cleavage Reagent Data

❖ To filter cleavage reagent data

1. Click the funnel icon, , next to the header of the column.
2. Select one of the following:
   - All: Returns the filtered search results to the results that were first loaded.
   - Custom: Opens the Custom Filter dialog box, shown in Figure 131.

     Figure 131. Custom Filter dialog box

     For information about using this type of dialog box, see "Filtering Results with Row Filters" on page 393.
   - Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked.
   - NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked.

Interpreting Search Results

This chapter describes the reports that you can generate in the Proteome Discoverer application and how to interpret the search results.

Contents

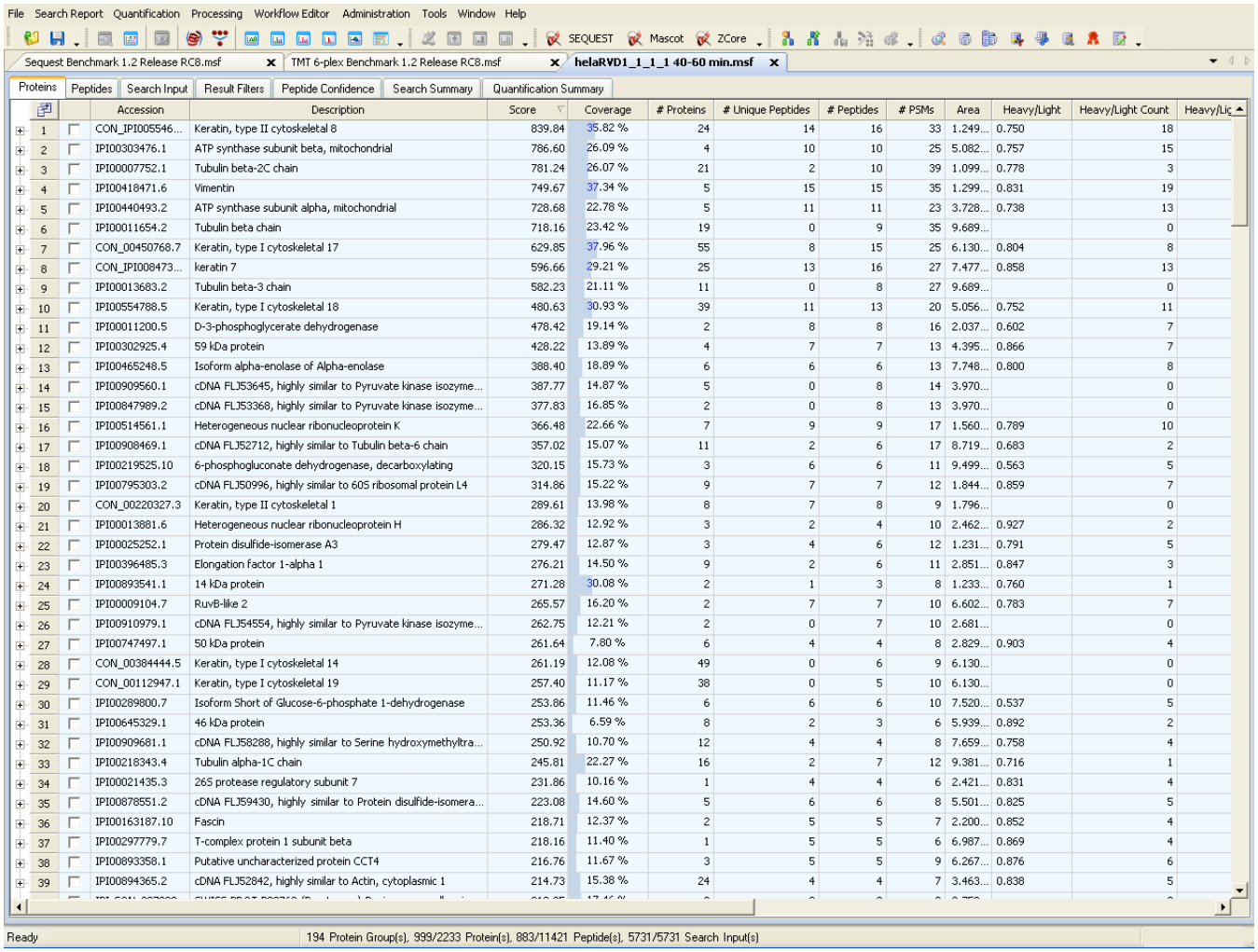
- Using the Results Report
- Using the Proteins Page
- Using the Peptides Page
- Using the Search Input Page
- Using the Result Filters Page
- Using the Peptide Confidence Page
- Using the Search Summary Page
- Using the Quantification Summary Page
- Using the Quan Spectra Page
- Creating a Report from Multiple Results

Using the Results Report

After the Proteome Discoverer application completes your search, it creates an MSF file, which contains the data and results from your search. This results report displays a list of matching peptides and proteins identified by the search engine that you specified. You can open this MSF file to access and interpret the search results. You can also display multiple search results within a single report.

The search results are displayed in multiple labeled pages of descriptive information in the application results window, as shown in Figure 132, and most of the commands on the Search Report menu now become available.

Figure 132. Example of the results report (MSF file)



	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	Area	Heavy/Light	Heavy/Light Count
1	CON_IPI005546...	Keratin, type II cytoskeletal 8	839.84	35.82 %	24	14	16	33	1.249...	0.750	18
2	IPI00303476.1	ATP synthase subunit beta, mitochondrial	786.60	26.09 %	4	10	10	25	5.082...	0.757	15
3	IPI00007752.1	Tubulin beta-2C chain	781.24	26.07 %	21	2	10	39	1.099...	0.778	3
4	IPI00418471.6	Vimentin	749.67	37.34 %	5	15	15	35	1.299...	0.831	19
5	IPI00440493.2	ATP synthase subunit alpha, mitochondrial	728.68	22.78 %	5	11	11	23	3.728...	0.738	13
6	IPI00011654.2	Tubulin beta chain	718.16	23.42 %	19	0	9	35	9.689...		0
7	CON_00450768.7	Keratin, type I cytoskeletal 17	629.85	37.96 %	55	8	15	25	6.130...	0.804	8
8	CON_IPI008473...	keratin 7	596.66	29.21 %	25	13	16	27	7.477...	0.858	13
9	IPI00013683.2	Tubulin beta-3 chain	582.23	21.11 %	11	0	8	27	9.689...		0
10	IPI00554788.5	Keratin, type I cytoskeletal 18	480.63	30.93 %	39	11	13	20	5.056...	0.752	11
11	IPI00011200.5	D-3-phosphoglycerate dehydrogenase	478.42	19.14 %	2	8	8	16	2.037...	0.602	7
12	IPI00302925.4	59 kDa protein	428.22	13.89 %	4	7	7	13	4.395...	0.866	7
13	IPI00465248.5	Isoform alpha-enolase of Alpha-enolase	388.40	18.89 %	6	6	6	13	7.748...	0.800	8
14	IPI00909560.1	cDNA FLJ53645, highly similar to Pyruvate kinase isozyme...	387.77	14.87 %	5	0	8	14	3.970...		0
15	IPI00847989.2	cDNA FLJ53368, highly similar to Pyruvate kinase isozyme...	377.83	16.85 %	2	0	8	13	3.970...		0
16	IPI00514561.1	Heterogeneous nuclear ribonucleoprotein K	366.48	22.66 %	7	9	9	17	1.560...	0.789	10
17	IPI00908469.1	cDNA FLJ52712, highly similar to Tubulin beta-6 chain	357.02	15.07 %	11	2	6	17	8.719...	0.683	2
18	IPI00219525.10	6-phosphogluconate dehydrogenase, decarboxylating	320.15	15.73 %	3	6	6	11	9.499...	0.563	5
19	IPI00795303.2	cDNA FLJ50996, highly similar to 60S ribosomal protein L4	314.86	15.22 %	9	7	7	12	1.844...	0.859	7
20	CON_00220327.3	Keratin, type II cytoskeletal 1	289.61	13.98 %	8	7	8	9	1.796...		0
21	IPI00013881.6	Heterogeneous nuclear ribonucleoprotein H	286.32	12.92 %	3	2	4	10	2.462...	0.927	2
22	IPI00025252.1	Protein disulfide-isomerase A3	279.47	12.87 %	3	4	6	12	1.231...	0.791	5
23	IPI00396485.3	Elongation factor 1-alpha 1	276.21	14.50 %	9	2	6	11	2.851...	0.847	3
24	IPI00893541.1	14 kDa protein	271.28	30.08 %	2	1	3	8	1.233...	0.760	1
25	IPI00009104.7	RuvB-like 2	265.57	16.20 %	2	7	7	10	6.602...	0.783	7
26	IPI00910979.1	cDNA FLJ54554, highly similar to Pyruvate kinase isozyme...	262.75	12.21 %	2	0	7	10	2.681...		0
27	IPI00747497.1	50 kDa protein	261.64	7.80 %	6	4	4	8	2.829...	0.903	4
28	CON_00384444.5	Keratin, type I cytoskeletal 14	261.19	12.08 %	49	0	6	9	6.130...		0
29	CON_00112947.1	Keratin, type I cytoskeletal 19	257.40	11.17 %	38	0	5	10	6.130...		0
30	IPI00289800.7	Isoform Short of Glucose-6-phosphate 1-dehydrogenase	253.86	11.46 %	6	6	6	10	7.520...	0.537	5
31	IPI00645329.1	46 kDa protein	253.36	6.59 %	8	2	3	6	5.939...	0.892	2
32	IPI00909681.1	cDNA FLJ58288, highly similar to Serine hydroxymethyltra...	250.92	10.70 %	12	4	4	8	7.659...	0.758	4
33	IPI00218343.4	Tubulin alpha-1C chain	245.81	22.27 %	16	2	7	12	9.381...	0.716	1
34	IPI00021435.3	26S protease regulatory subunit 7	231.86	10.16 %	1	4	4	6	2.421...	0.831	4
35	IPI00878551.2	cDNA FLJ59430, highly similar to Protein disulfide-isomera...	223.08	14.60 %	5	6	6	8	5.501...	0.825	5
36	IPI00163187.10	Fascin	218.71	12.37 %	2	5	5	7	2.200...	0.852	4
37	IPI00297779.7	T-complex protein 1 subunit beta	218.16	11.40 %	1	5	5	6	6.987...	0.869	4
38	IPI00893358.1	Putative uncharacterized protein CCT4	216.76	11.67 %	3	5	5	9	6.267...	0.876	6
39	IPI00894365.2	cDNA FLJ52842, highly similar to Actin, cytoplasmic 1	214.73	15.38 %	24	4	4	7	3.463...	0.838	5

The labeled pages include the following:

- Proteins page: See "Using the Proteins Page" on page 237.
- Peptides page: See "Using the Peptides Page" on page 288.
- Search Input page: See "Using the Search Input Page" on page 346.
- Result Filters page: See "Using the Result Filters Page" on page 349.
- Peptide Confidence page: See "Using the Peptide Confidence Page" on page 349.
- Search Summary page: See "Using the Search Summary Page" on page 349.
- Quantification Summary page: See "Summarizing the Quantification" on page 523. The Proteome Discoverer application generates this page only if you included a quantification node in your workflow.
- Quan Spectra page: See "Displaying Quantification Spectra" on page 527. The Proteome Discoverer application generates this page only if you included a reporter ion quantification node in your workflow.

You can also match likely proteins to your analyzed data and access isotope, chromatogram, and other graphical views. The views provide detailed graphical information about the selected peptides. You can display more than one view to perform a comparative analysis of your selected peptide or proteins. Use the shortcut menus on the pages of the results report to alter the details in your report or to copy information. To customize your report, you can move the columns to a different location in the report. You can also quickly sort your columns by clicking the column header and set different types of filtering to sort your search results.

Opening MSF Files

You can open an MSF (results report) file through the application menu bar, toolbar, or job queue. You can optionally set filters to reduce the number of peptides and proteins to load.

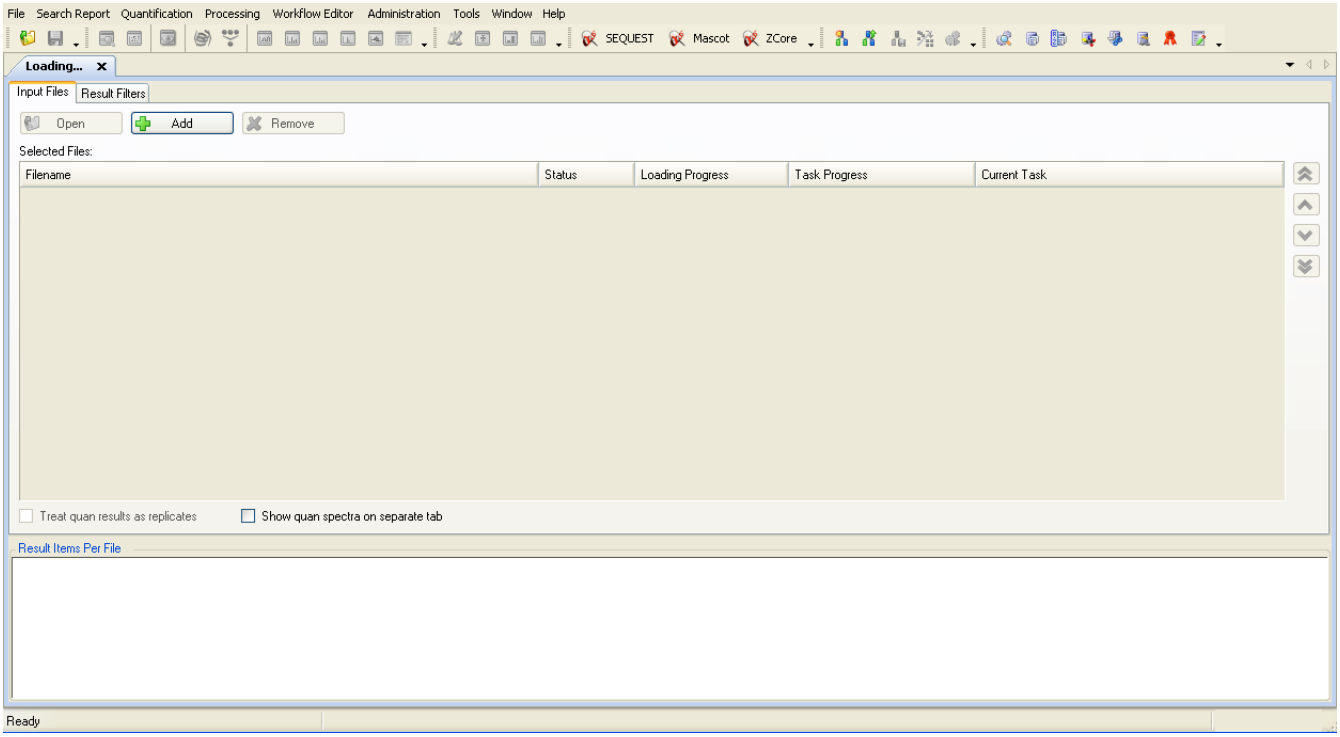
- To open or load an MSF (report) file from the menu bar or toolbar
- To filter the data in the MSF file before opening it
- To open an MSF report from the job queue

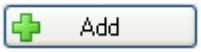
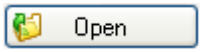
❖ To open or load an MSF (report) file from the menu bar or toolbar

1. In the Proteome Discoverer window, choose **File > Open Report**, click the **Open Report** icon,  , or press CTRL+O.

   The Input Files page appears, as shown in Figure 133.

Figure 133. Input Files page



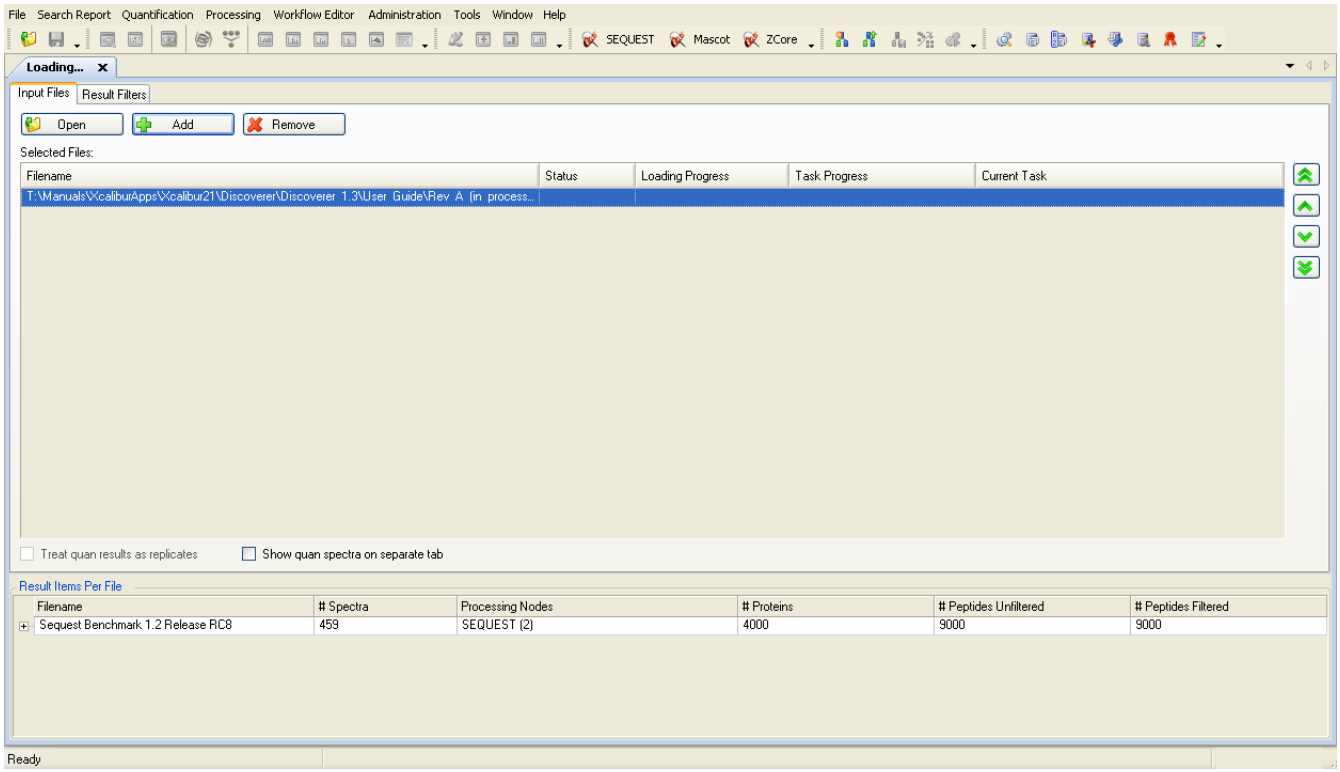
2. Click the **Add** icon,  .
3. In the Add Analysis File(s) dialog box, browse to the MSF file or files that you want to load, and click **Open**.
4. When the files are updated, click the **Open** icon,  .

   The application loads the files in the order in which they are sorted in the Selected Files area. It displays the loading status for each file:

   - Empty: The file is not loaded.
   - Loading: The file is currently being loaded.
   - Green check mark: The file has been successfully loaded.

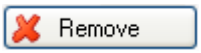
   The path and file names now appear on the Input Files page, as shown in Figure 134. The file names appear in the Selected Files area and below it in the Result Items Per File area.

Figure 134. Input Files page showing selected file to open



5. To open more files, click **Add** again.

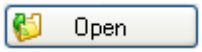
   The added file or files now appear in the Selected Files area.

6. To remove a file from the Selected Files area, select the file and click the **Remove** icon,  .

7. Use the following icons to rearrange the files listed in the Selected Files area, if necessary:
   -  : Moves the selected file in the Selected Files area to the top of the list.
   -  : Moves the selected file in the Selected Files area one line above its present location.
   -  : Moves the selected file in the Selected Files area one line below its present location.
   -  : Moves the selected file in the Selected Files area to the bottom of the list.
8. When you are opening multiple MSF files containing quantification results and want to treat protein-level quantification values with the same ratio names and the same quantification method as replicates (that is, the protein ratios of the individual files are averaged into a replicate ratio), select the **Treat Quan Results as Replicates** option. When you want to change the setting after the report is open, you can use the Treat Quan Results as Replicate option on the General page of the Quantification Method Editor dialog box. For more information about setting up the quantification method for multiple MSF files, see "Setting Up the Quantification Method for Multiple Input Files" on page 503.
9. To display the intensities and ratios for all spectra, regardless of whether they have been identified in reporter ion quantification, select the **Show Quan Spectra on Separate Tab** check box.

   This option generates the Quan Spectra page in the MSF report only if you included a reporter ion quantification node in your workflow. For more information on this page, see "Displaying Quantification Spectra" on page 527. You might want to use this option to highlight spectra that are unidentified but have significant ratios for the reporter ions. The presence of these spectra indicates that either a peptide was not in the database or that the peptide itself has a post-translational modification. These spectra indicate that you should try another search of the data with different search parameters.

   The Result Items Per File line is read-only.

   To filter out some of the peptides before you open the file, perform the steps given in "To filter the data in the MSF file before opening it."
10. If you do not want to filter the results in the MSF file before you open it, click the **Open** icon, 

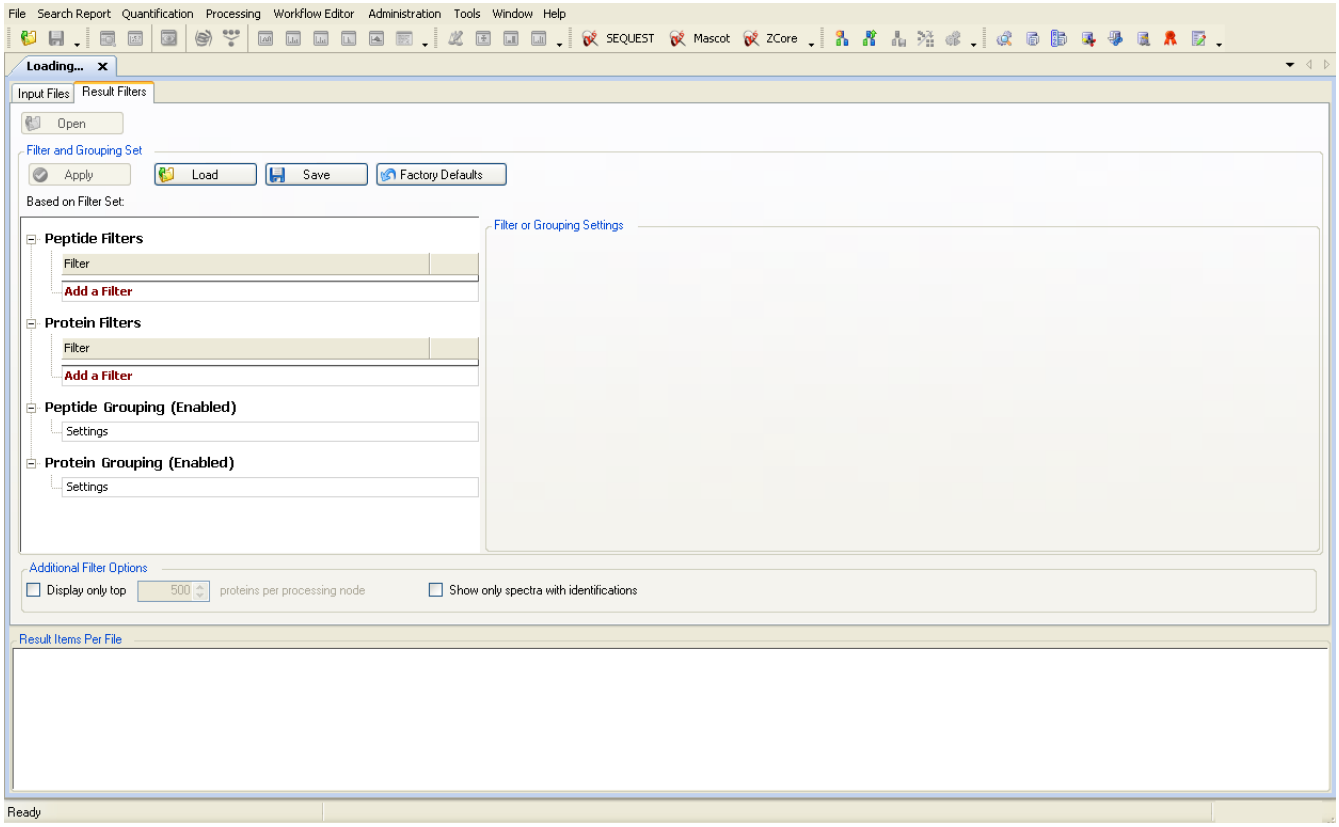
.

    When the Proteome Discoverer application needs to update the file as it is being loaded, you will see "Updating" in the Status box. To interrupt the update process, double-click **Updating** and click **OK** in the confirmation box that appears.

❖ To filter the data in the MSF file before opening it

1. Click the **Result Filters** tab, shown in Figure 135.

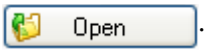
Figure 135. Result Filters page



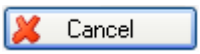
2. For instructions on setting the parameters on this page, see "Filtering Data" on page 359. The Filter and Protein Grouping Set Handling, Filter and Protein Grouping Set, Peptide Filters, and Filter and Protein Grouping (Enabled) areas of the dialog box are identical to those on the Result Filters page in the MSF file.

3. To limit the number of proteins displayed in the MSF file, select the **Display Only Top** check box in the Additional Filter Options area, and indicate the number of highest-scoring proteins that you want to load in the Proteins Per Processing Node check box.

4. To load only spectra with an identified peptide, select the **Show Only Spectra With Identifications** check box.

5. Click the **Open** icon, 

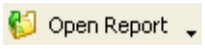
.

   To cancel the loading of the MSF file, click the **Cancel** icon, 

.

   If you do not cancel the loading, the MSF file now opens with the Proteins page displayed by default.

❖ **To open an MSF report from the job queue**

1. Open the job queue by choosing **Administration > Show Job Queue.**
2. In the job queue, click the Completed row for the job that you are interested in.
3. Click the **Open Report** icon,  .

   The Input Files page appears, as shown in Figure 133 on page 215. Follow the procedure given in "To open or load an MSF (report) file from the menu bar or toolbar" on page 215.

Input Files Page Parameters

Table 37 describes the parameters on the Input Files page.

Table 37. Input Files page parameters (Sheet 1 of 3)

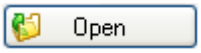
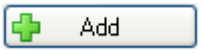
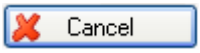
Parameter	Description
	Opens the MSF file (results report).
	Opens the Add Analysis File(s) dialog box so you can browse for the file or files to open.
	Removes the selected files from the Selected Files area.
	Cancels the loading of the selected MSF file or files.
Selected Files	Lists the paths and names of the files that you want to open.
Filename	Displays the path and name of the selected MSF file.
Status	Displays the status of the loading after you click Open: • Empty: The file is not loaded. • Loading: The file is currently being loaded. • Green check mark: The file has been successfully loaded.
Loading Progress	Displays a meter bar showing what percentage of the file is currently loaded after you click Open.
Task Progress	Displays a meter bar showing what percentage of the current task is completed after you click Open.

Table 37. Input Files page parameters (Sheet 2 of 3)

Parameter	Description
Current Task	Displays the part of the MSF report that the Proteome Discoverer application is currently loading after you click Open, for example, loading peptides.
	Moves the selected file in the Selected Files area to the top of the list.
	Moves the selected file in the Selected Files area one line above its present location.
	Moves the selected file in the Selected Files area one line below its present location.
	Moves the selected file in the Selected Files area to the bottom of the list.
Treat Quan Results as Replicates	Treats protein-level quantification values with the same ratio names and the same quantification method as replicates (that is, the protein ratios of the individual files are averaged into a replicate ratio).
Show Quan Spectra on Separate Tab	Displays the intensities and ratios for all spectra on the Quan Spectra page in the MSF file, regardless of whether they have been identified in reporter ion quantification. For more information on the Quan Spectra page, see "Using the Quan Spectra Page" on page 351.
Result Items Per File	Displays information about each MSF file selected in the Selected Files area.
Filename	Displays the name of the MSF file.
# Spectra	Displays the total number of spectra in each MSF file to load.
Processing Nodes	Displays the name and node number of the search engine that produced the results in the MSF file: SEQUEST, Mascot, or ZCore. If more than one processing node created the peptides and proteins in one file, the data for each processing node is separated by commas.
# Proteins	Displays the number of proteins for each processing node.

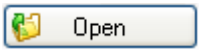
Table 37. Input Files page parameters (Sheet 3 of 3)

Parameter	Description
# Peptides Unfiltered	Displays the total number of peptides in the MSF file before any filters were applied.
# Peptides Filtered	Displays an estimate of the number of peptides in the MSF file that are retained after the filters are applied.

Result Filters Page Parameters

Table 38 lists the parameters on the Result Filters page. For information about the parameters on this page, see "Result Filters Page Parameters" on page 361. The Filter and Protein Grouping Set Handling, Filter and Protein Grouping Set, and Filters and Protein Grouping Settings areas of the dialog box are the same as those on the Result Filters page in the MSF file.

Table 38. Result Filters page parameters

Feature	Description
	Opens the selected MSF file or files.
Additional Filter Options	Determines how many proteins are displayed in the MSF file or files.
Display Only Top	Determines whether the number of highest-scoring proteins is displayed in the MSF file.
Show Only Spectra With Identifications	Loads only spectra with an identified peptide.
Result Items Per File	The parameters in this area are the same as those in the Result Items Per File area on the Input Files page. For a description of these parameters, see Table 37.

Saving MSF Files

The Proteome Discoverer application does not automatically save your changes to reports or filter settings. You must save your changes to keep them for future use.

❖ **To save a report in an MSF file**

- Choose **File > Save Report**.

The changes or filter settings are stored in the results (MSF) file.

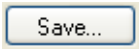
Closing MSF Files

You can close an individual MSF file or close all views and MSF files and exit the application.

❖ To close a report in an MSF file without saving changes

1. In the Proteome Discoverer window, click the report page that you want to close.
2. Choose **File > Close**, or right-click the report that you want to close, and choose **Close** from the shortcut menu.

❖ To exit and close all views and reports

1. Choose **File > Save Report** to save your changes.
2. Choose **Workflow > Save as Template** to save your workflow.
3. Click the **Result Filters** page. Click 

to save your filter settings.
4. Choose **File > Exit**.

Importing SRF Files

To import SRF files created in the BioWorks 3.3.1 application, which preceded the Proteome Discoverer application, see "Migrating from BioWorks to the Proteome Discoverer Application" on page 715. It gives you instructions for using the SRF File Import Wizard dialog box activated by choosing **File > Import Search Results from SRF**.

Organizing Reports

In the initial results report, you can customize the data by using these features:

- Selecting the Columns to Display
- Filtering Data
- Specifying a Mass Tolerance Unit
- Displaying Peptides
- Grouping Data on the Proteins and Peptides Pages
- Sorting Columns
- Assigning Index Numbers to Rows
- Selecting Peptides and Proteins
- Creating and Applying Custom Layouts
- Exporting Report Data

Selecting the Columns to Display

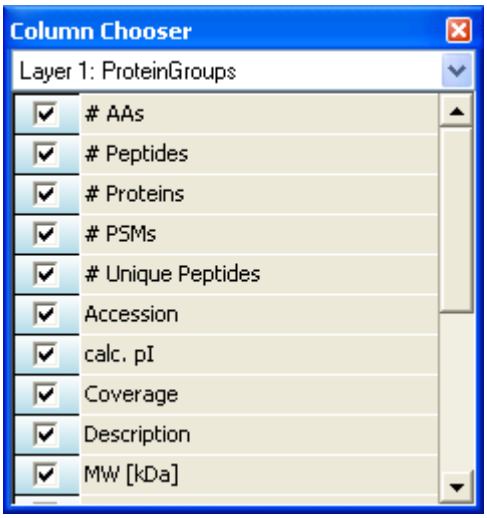
You can choose which columns to display on the Proteins, Peptides, and Search Input pages.

❖ To select the columns to display

1. In the upper left corner of each page, click the **Column Chooser** icon, .

   The Column Chooser dialog box appears. Figure 136 gives an example of this dialog box for the Proteins page.

   Figure 136. Column Chooser dialog box

2. To display a column, select the check box next to the column name in the dialog box. To hide a column, clear the selected check box next to the column name.

   The Proteome Discoverer application instantly makes the selected columns visible and the cleared columns invisible.

   You can also remove columns by dragging the column title above the column-title line.

For a description of the columns available on the Proteins page, see "Proteins Page Parameters" on page 241. For a description of the columns available on the Peptides page, see "Peptides Page Parameters" on page 291. For a description of the columns available on the Search Input page, see "Search Input Page Parameters" on page 347.

Filtering Data

For information about filtering the search results on the Proteins, Peptides, and Search Input pages, see "Filtering Data" on page 359.

Specifying a Mass Tolerance Unit

You can change the units that the Proteome Discoverer application uses to report the values in the ΔM [ppm] column on the Peptides page. This column displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion. By default, the application reports these values in parts per million (ppm).

❖ To change the mass tolerance unit

1. Right-click the Proteins, Peptides, or Search Input page and choose **Mass Tolerance Unit** from the shortcut menu.
2. Select one of the following units:
   - Da: Reports values in daltons.
   - ppm: Reports values in parts per million.
   - mmu: Reports values in milli-mass units.

Displaying Peptides

You can display just the highest-scoring peptide among all peptides identified with the same sequence for a given protein in the search, regardless of the node type, node, activation type, or mass spectrometer used in the analysis and search.

You might want to identify the same peptide sequence multiple times for several possible reasons:

- You triggered several MS/MS spectra for the same precursor, probably from different charge states.
- You identified the same spectrum with multiple search engines, for example, SEQUEST and Mascot.
- You identified the same peptide sequence from MS/MS spectra with different activation types.
- You identified the same peptide sequence in different modification states, for example, phosphorylation at different positions.

The highest-scoring peptide is defined by the search node itself. Each node can have several peptide scores, and one of these is defined as the main score:

- For SEQUEST, it is the peptide with the highest cross-correlation (XCorr) number.
- For ZCore, it is the peptide with the highest probability.
- For Mascot, it is the peptide with the highest ion score.

Go to the following sections:

- To display the peptides included in a protein from the Proteins page
- To immediately display a peptide on the Search Input page

❖ To display the peptides included in a protein from the Proteins page

1. On the Proteins page, click + to the left of the appropriate protein to display its peptides.
2. Right-click the appropriate peptide and choose **Show in Peptide View** from the shortcut menu.

   The peptide is now highlighted on the Peptides page.
3. On the Peptides page, click + to the left of a peptide on the first level to display the protein that a particular peptide belongs to.
4. Click + to the left of a protein on the second level to display the other peptides that belong to this same protein.

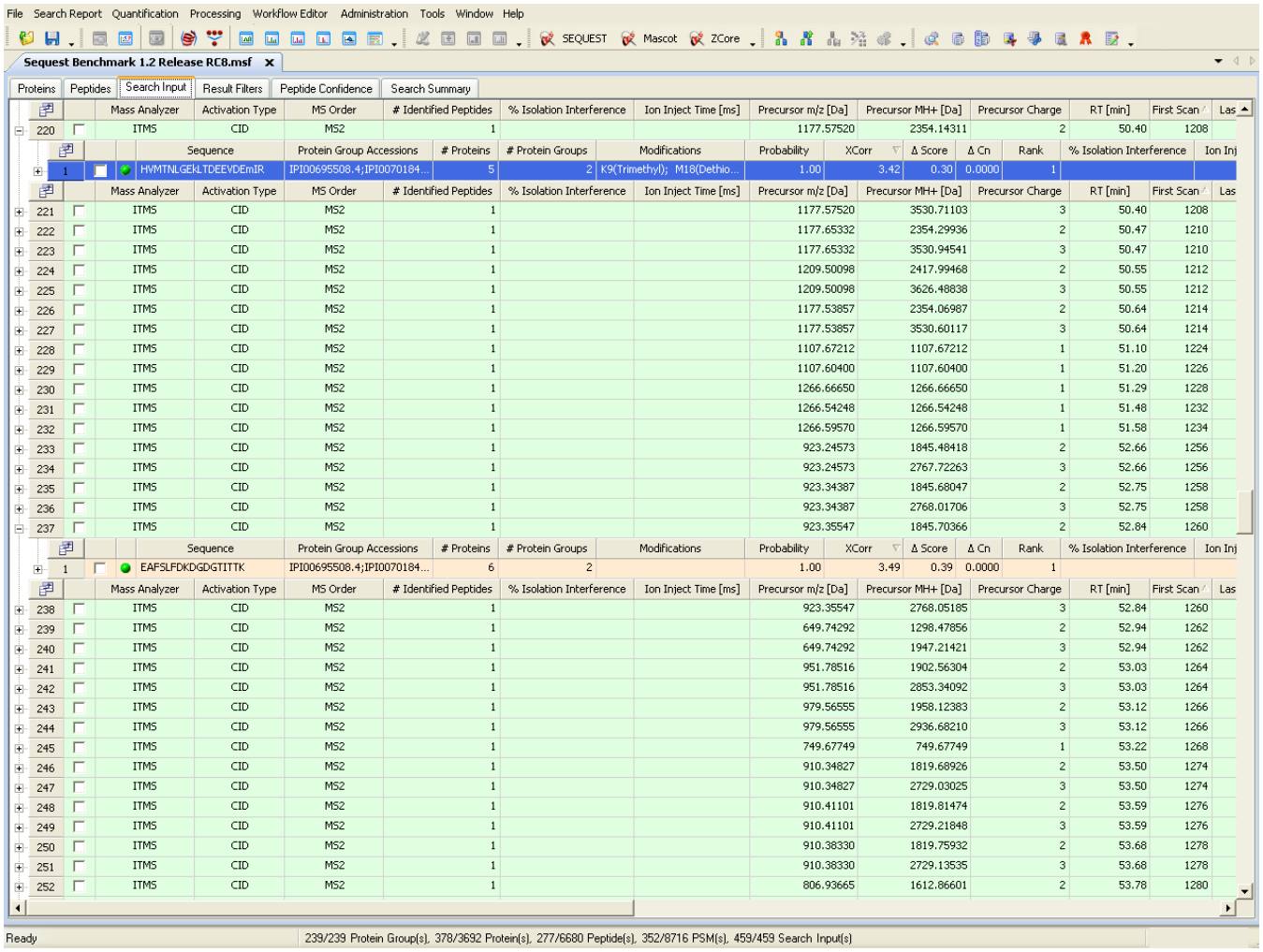
When you click a peptide on the Proteins or Peptides page, the Proteome Discoverer application can immediately highlight it on the Search Input page so that you can view detailed information about the spectra and see how many identified peptides were found and the proteins that the peptide belongs to.

❖ To immediately display a peptide on the Search Input page

1. On the Proteins or Peptides page, ungroup peptides, select a peptide, and choose **Search Report > Show Peptides Group Members View**.
2. Right-click the peptide, and choose **Show in Search Input View** from the shortcut menu.

   The Search Input page opens and highlights the peptide of interest, as shown in Figure 137. It also displays any other peptides grouped with that scan number.

Figure 137. Peptide shown on the Search Input page



Grouping Data on the Proteins and Peptides Pages

By default, the Proteome Discoverer application groups together the most prevalent protein and its variations. It also groups peptides together. However, you can remove this grouping. For more information on grouping and ungrouping proteins and peptides, see "Grouping Proteins" on page 401 and "Grouping Peptides" on page 413, respectively.

On the Proteins page, you can group data by column header, and on the Peptides page, you can group data by column header or by sequence.

- To group data by column header
- To group peptides by sequence

❖ To group data by column header

1. Right-click the Peptides or Proteins page, and choose **Show 'Group by Column' Panel** from the shortcut menu.

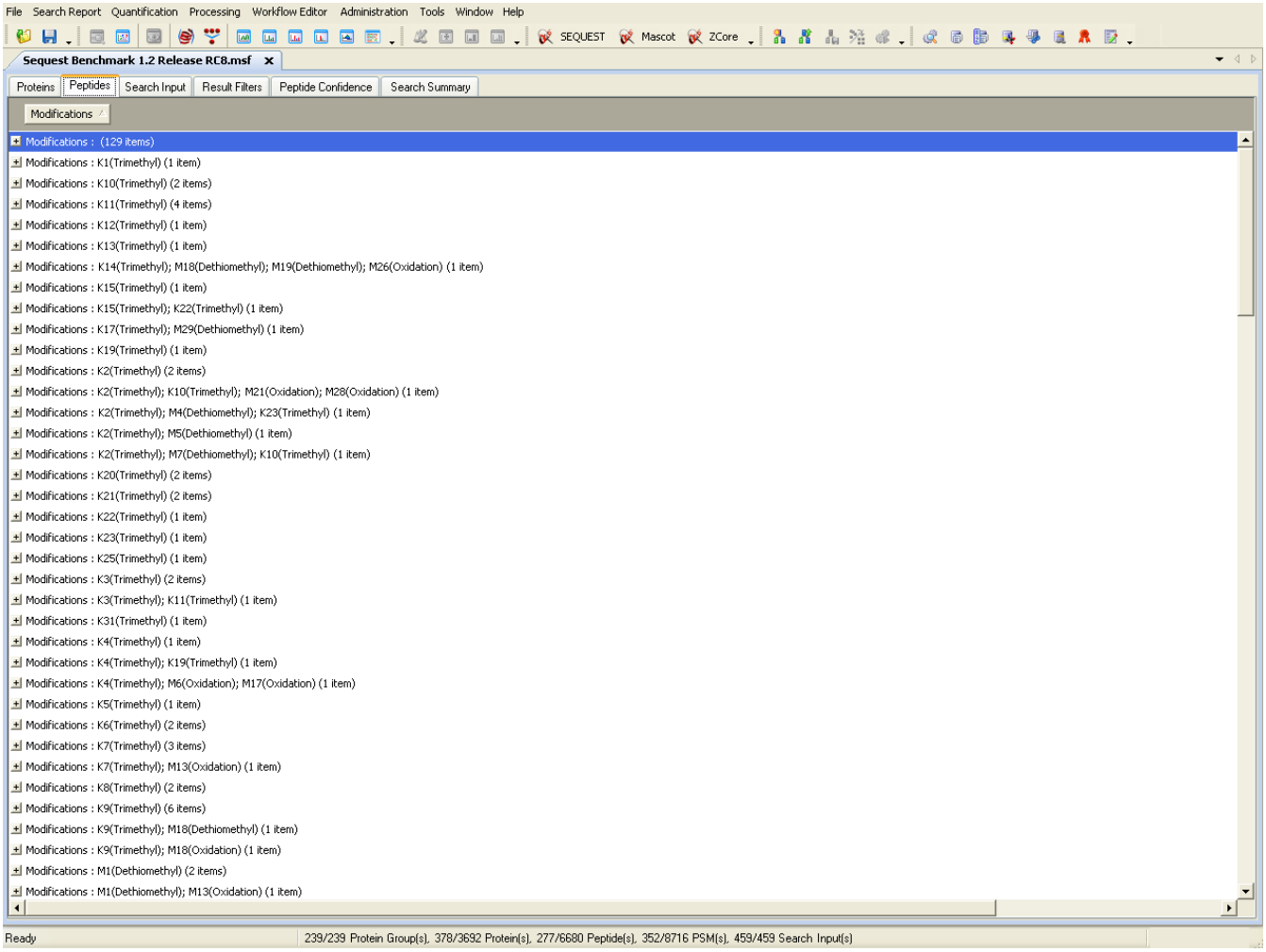
   The following message appears in the upper left corner of the Discoverer window:

   Drag a column header here to group by that column.

2. Drag a column header over the message.

   The information about the page is grouped by the selected column header. For example, in Figure 138, the Modifications column header was dragged over the message, so the peptides on the Peptides page are grouped by modifications.

Figure 138. Peptides grouped by Modifications column header



On the Peptides page, you can group together peptide identifications with the same underlying sequence under the same collapsible heading. The individual identifications within the group might differ in detected modifications, activation type, and particular search engine results. Use this feature when viewing results from multiple searches in a single report. Entries within the peptide group should still display no duplicate references.

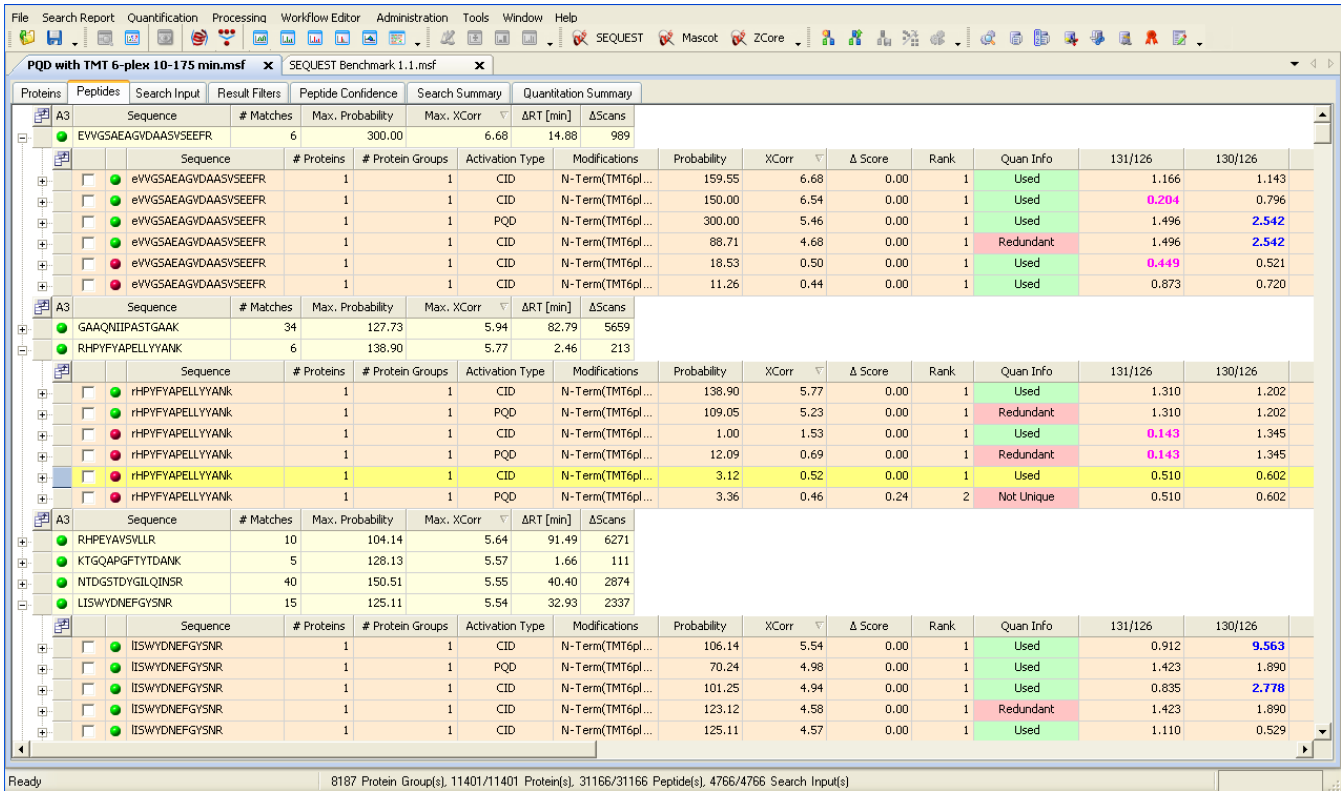
❖ To group peptides by sequence

1. Right-click the Peptides page, and choose **Show Peptide Groups** from the shortcut menu.

   The page displays the peptides in a summary-type view, as shown in Figure on page 291, showing only the highest-scoring peptides.

2. Click + to the left of a peptide of interest to see the other peptides that contain the same sequence, as shown in Figure 139.

Figure 139. Peptides grouped by sequence



Sorting Columns

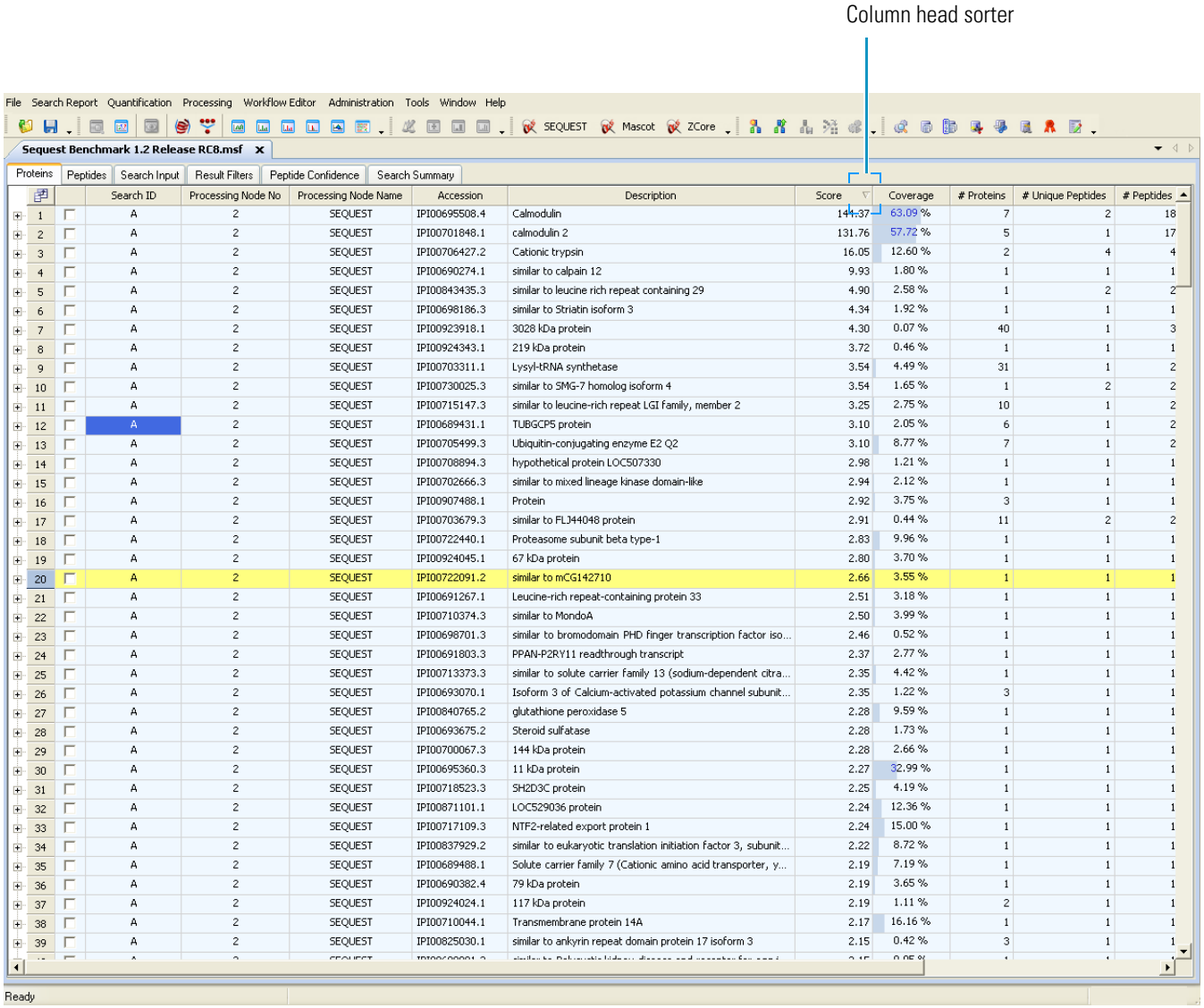
You can sort a column's entries in ascending or descending order. The small triangle to the right of the column header shows if the column is in ascending ( ) or descending ( ) order.

❖ To sort a column

1. Open the search results.
2. In the header of the column whose entries you want to sort, click to sort in ascending order or to sort in descending order.

   Figure 140 shows an example of data sorted in descending order in the Score column.

Figure 140. Column header sorting example



Assigning Index Numbers to Rows

You can assign each row on the Proteins, Peptides, or Search Input page an automatically assigned index number to facilitate identification of the rows. You can add row index numbers to all rows, whether visible or invisible, or just add them to visible rows. You can also specify that the assigned numbers remain with the rows even after sorting or filtering operations.

- To add row index numbers to visible rows only
- To add row index numbers to both visible and invisible rows
- To add fixed row index numbers
- To remove row index numbers

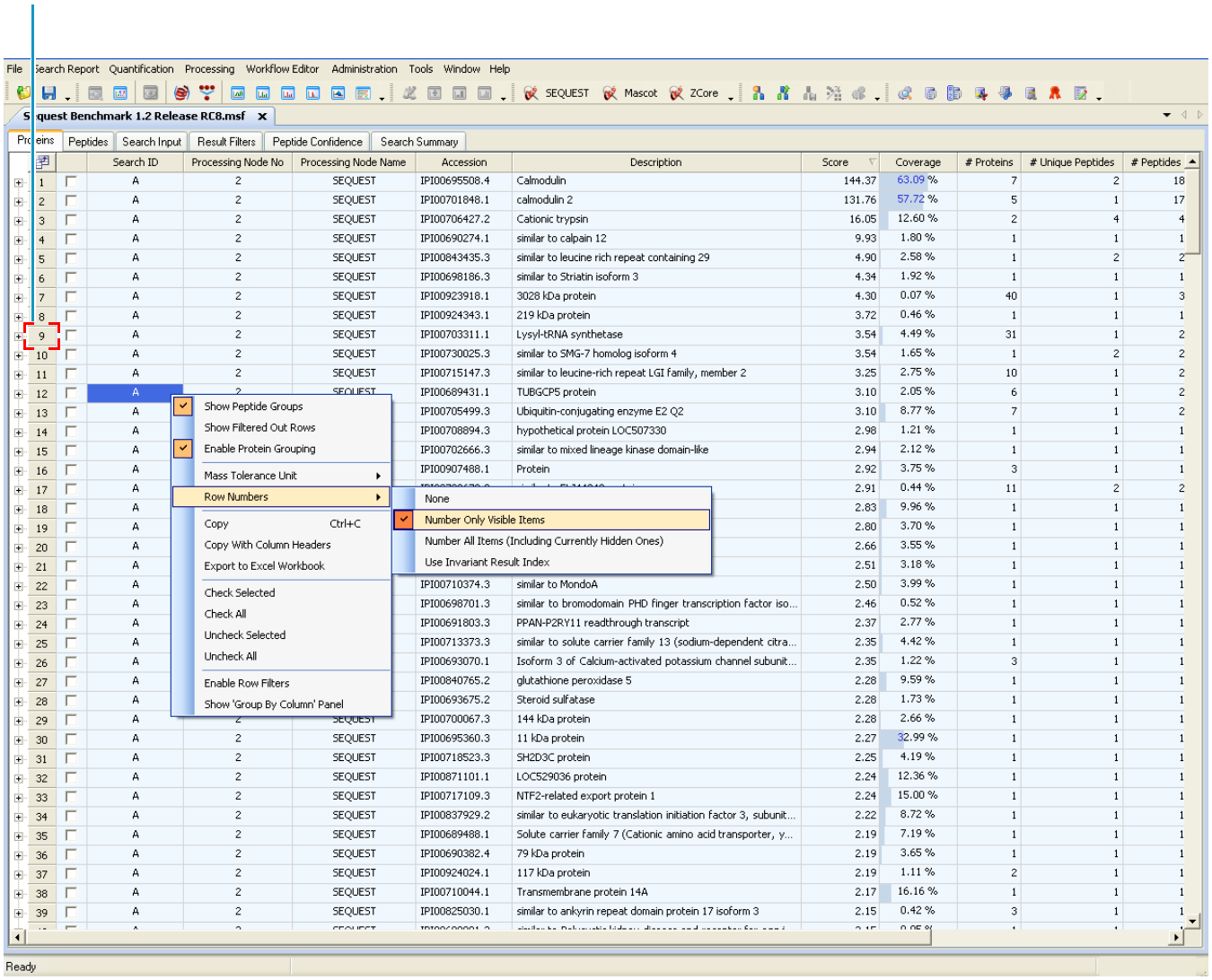
❖ To add row index numbers to visible rows only

1. Right-click the Peptides, Proteins, or Search Input page to display the shortcut menu.

   A list of options is displayed, as shown in Figure 141.

2. Choose **Row Numbers > Number Only Visible Items**.

Figure 141. Row-numbering options



❖ To add row index numbers to both visible and invisible rows

1. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
2. Choose **Row Numbers > Number All Items (Including Currently Hidden Ones)**.

The Proteome Discoverer application assigns a sequential number to each row from the top row to the bottom row.

❖ To add fixed row index numbers

1. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
2. Choose **Row Numbers > Use Invariant Result Index**.

❖ To remove row index numbers

1. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
2. Choose **Row Numbers > None**.

Selecting Peptides and Proteins

On the Proteins, Peptides, and Search Input pages, the Proteome Discoverer application has two selection mechanisms that you can use to select peptides and proteins. When you want to select one or more peptides or proteins for display purposes, such as generating one of the graphs on the Search Report menu, left-click the cursor on the appropriate peptides or proteins to select them.

You can also select several or all of a particular row type, such as proteins or peptides, and press the space bar to select them.

When you want to export peptides or proteins to a FASTA or Excel file, you must select the appropriate items with a check mark in the  box. You can mark all peptides or proteins in the MSF report or just the selected ones.

❖ To select all proteins or peptides in the MSF report

1. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
2. Choose **Check All**.

❖ To remove all proteins or peptides in the MSF report

1. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
2. Choose **Uncheck All**.

❖ To select specific proteins or peptides in the MSF report

1. Highlight the peptides or proteins.
2. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
3. Choose **Check Selected**.

❖ To remove a specific proteins or peptides from the MSF report

1. Highlight the peptides or proteins.
2. Right-click the Proteins, Peptides, or Search Input page to display the shortcut menu.
3. Choose **Uncheck Selected**.

Creating and Applying Custom Layouts

On the Proteins, Peptides, and Search Input pages, you can create a layout of a report page, save the layout changes, and apply the saved layout to other results reports or make it the default layout. After you have used the Column Chooser options, you can save your changes to a default layout or a special layout for your results report.

You can save two types of layouts, multiconsensus and single report. For more information about multiconsensus reports, see "Finding Common or Unique Proteins in Multiple Searches" on page 353.

Custom layout settings have the following limitations:

- When you make a change to a layout in a report created by one type of search engine and save that layout as the default and then apply it to a report created by another type of search engine, the change may not appear in the latter report. For example, layout settings saved in a SEQUEST search report behave differently for a Mascot search report, especially when used as the default layout. On the Peptides page, SEQUEST by default sorts by the XCorr value in descending order, ZCore by default sorts by the Probability value in descending order, and Mascot by default sorts by the IonScore value in descending order. When you apply this default layout in a SEQUEST-generated report to a Mascot-generated report, the XCorr value does not appear in the Mascot-generated report, because Mascot reports do not have an XCorr column; they have an IonScore column. The changes in a default layout made to a report created by a certain search engine can only reliably be applied to reports generated by the same search engine.
- When you generate a multiconsensus report by using more than one search engine, the results report by default includes at least two of the following columns:
  - XCorr column (SEQUEST)
  - Probability column (ZCore)
  - IonScore column (Mascot)

  If you make a change to this default layout but want to return to it later, choose **Search Report > Layout > Restore Factory Default**.
- You cannot apply a layout for a multiconsensus report to a single report, nor can you apply a layout for a single report to a multiconsensus report.

Go to the following sections:

- To create and save a layout
- To apply a new layout to a results report
- To create a default layout
- To apply the default layout to your report

- To return to the layouts in effect when you first installed the Proteome Discoverer application
- To save a results report layout

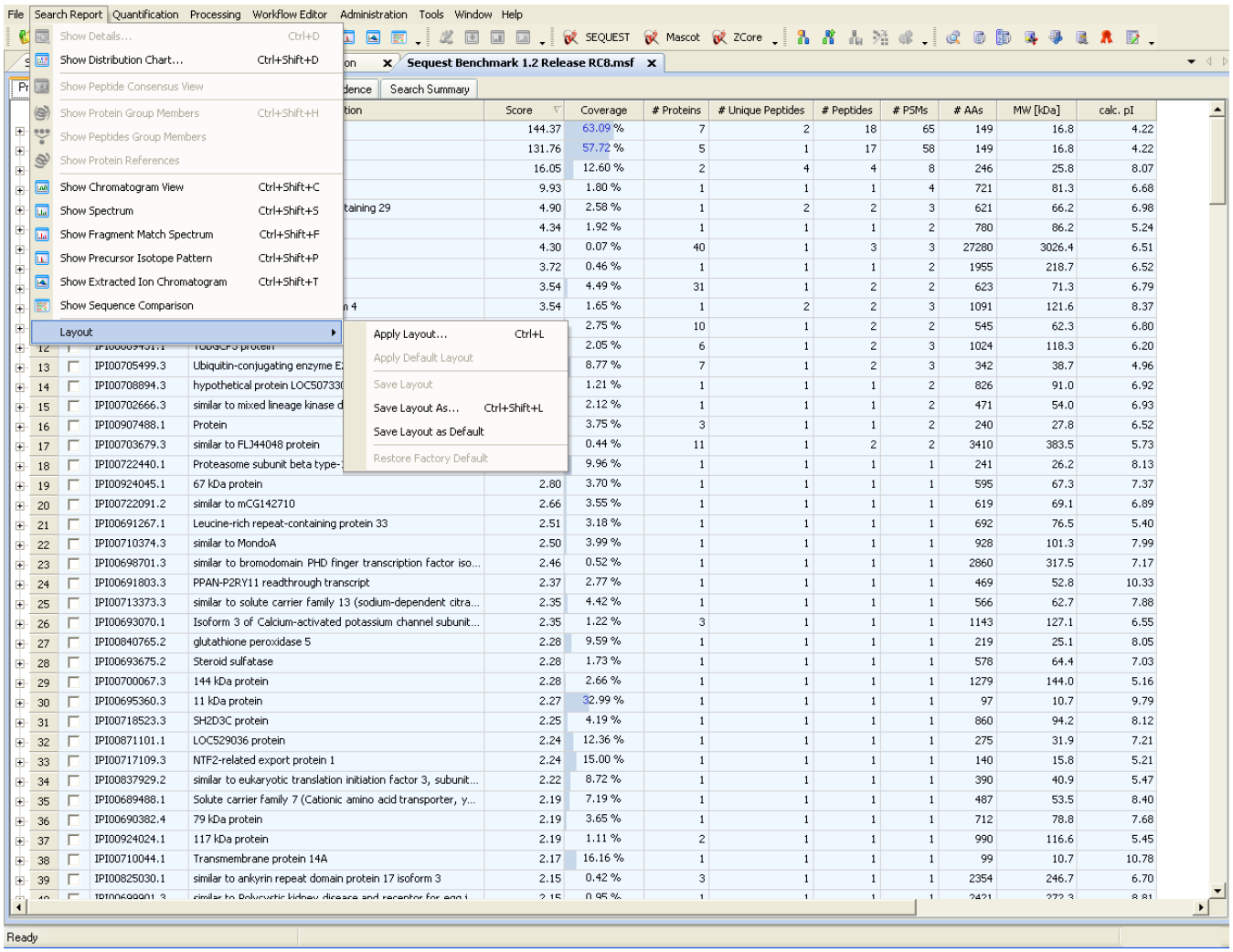
❖ To create and save a layout

1. Create a layout on the Proteins, Peptides, or Search Input page.

   See "Selecting the Columns to Display" on page 223 and "Grouping Data on the Proteins and Peptides Pages" on page 226.

2. Choose **Search Report > Layout > Save Layout As** to save the column and row changes, as shown in Figure 142.

Figure 142. Layout menu options



3. In the Save As dialog box, browse to the file in which you want to save the layout and select it, or type the path and name of the file in the File Name box.

   The file must have a .dlyt suffix for a single report. The file must have an .mlyt suffix for a multiconsensus report.

4. Click **Save** to save the layout with the specified file name.

   The view closes and your layout properties are stored in the specified .dlyt or .mlyt file in the layout folder.

❖ To apply a new layout to a results report

1. Click the page of the results report that you want to apply the new layout to.
2. Choose **Search Report > Layout > Apply Layout**.
3. In the Open dialog box, browse to the layout that you want to apply and select it.
4. Click **Open** to apply the layout to your report.

   The properties of the selected layout are applied to the results report.

❖ To create a default layout

1. Create a layout on the Proteins, Peptides, or Search Input page.

   See "Selecting the Columns to Display" on page 223 and "Grouping Data on the Proteins and Peptides Pages" on page 226.

2. Choose **Search Report > Layout > Save Layout As Default** to save the column and row configuration.

   The file name must have a .dlyt suffix for a single report. The file name must have an .mlyt suffix for a multiconsensus report.

You can apply the row and column properties to any results report. To apply the default layout to any report, choose Search Report > Apply Default Layout. If a report is already in the default format, this command is not available on the Layout menu.

❖ To apply the default layout to your report

1. Click the appropriate page of the results report.
2. Choose **Search Report > Layout > Apply Default Layout**.

   You must have modified the layout of the page before this command becomes available on the Layout menu.

   The properties of the default layout are applied to the results report.

❖ **To return to the layouts in effect when you first installed the Proteome Discoverer application**

1. Click the appropriate page of the results report.
2. Choose **Search Report > Layout > Restore Factory Default**.

❖ **To save a results report layout**

1. Open a search results report.
2. Modify the appropriate page:
   - Make changes to the layout, choose **Search Report > Layout > Save Layout As** to save them in a new layout, and then make additional changes to the new layout.

     –or–

   - Choose **Search Report > Layout > Apply Layout** to apply a different layout to the appropriate page.
3. Choose **Search Report > Layout > Save Layout**.

Exporting Report Data

You can copy columns of data from the Proteins, Peptides, and Search Input pages so that you can use them in presentations or in Excel spreadsheets. For instructions on the procedure for exporting data to an Excel spreadsheet, see "Exporting Search Results to Excel" on page 604.

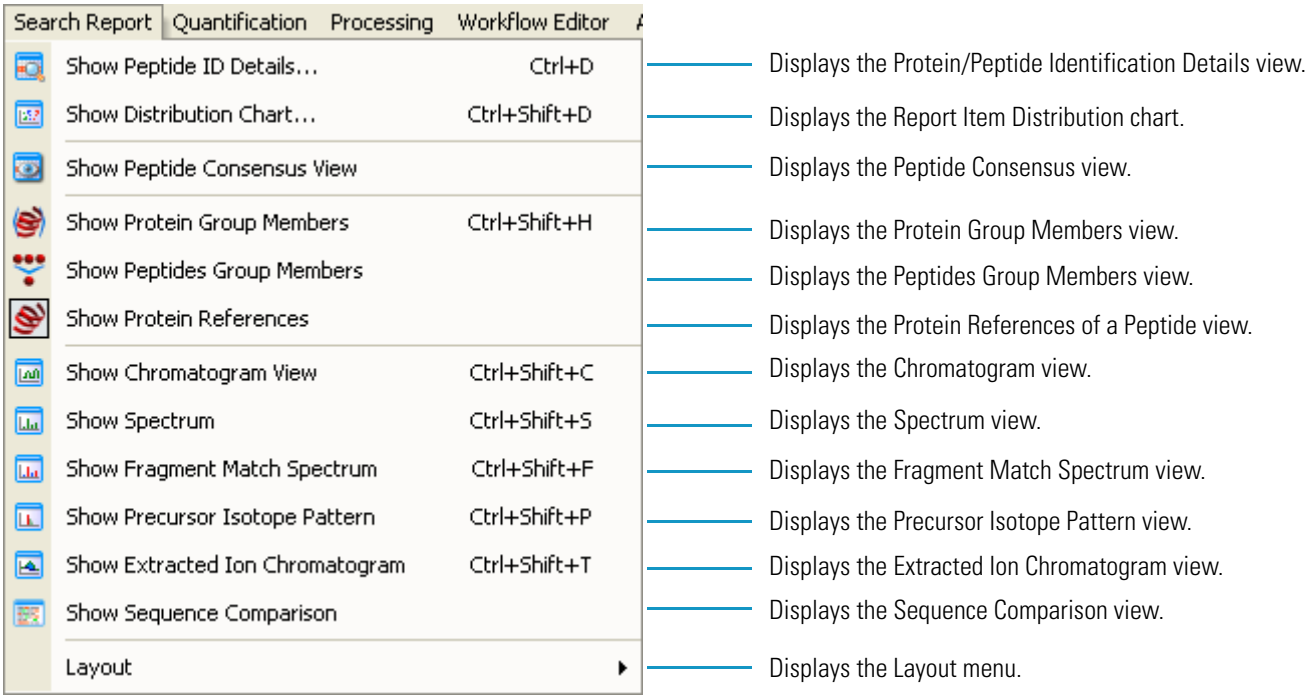
Controlling the Search Report Windows

For information about moving the search report windows that appear when you select commands on the Search Report menu, see "Search Results Window Icons" on page 33.

Using the Search Report Menu

You can use the commands on the Search Report menu, shown in Figure 143, with any of the Proteome Discoverer application's report pages. The following sections describe these report pages in detail.

Figure 143. Search Report menu commands



Using the Proteins Page

The Proteins page lists all the identified proteins and the associated peptides found in a sample during a search. You can examine the search results in terms of protein identification, as well as access more details about the peptide identifications and corresponding information from the search input. The Proteins page gives you detailed tabular information, a shortcut menu, and access to the peptide information.

- Displaying the Proteins Page
- Interpreting Your Results with the Protein Identification Details View
- Interpreting Your Results with the Report Item Distribution Chart
- Interpreting Your Results with the Sequence Comparison View
- Interpreting Your Results with the Protein Group Members View

Displaying the Proteins Page

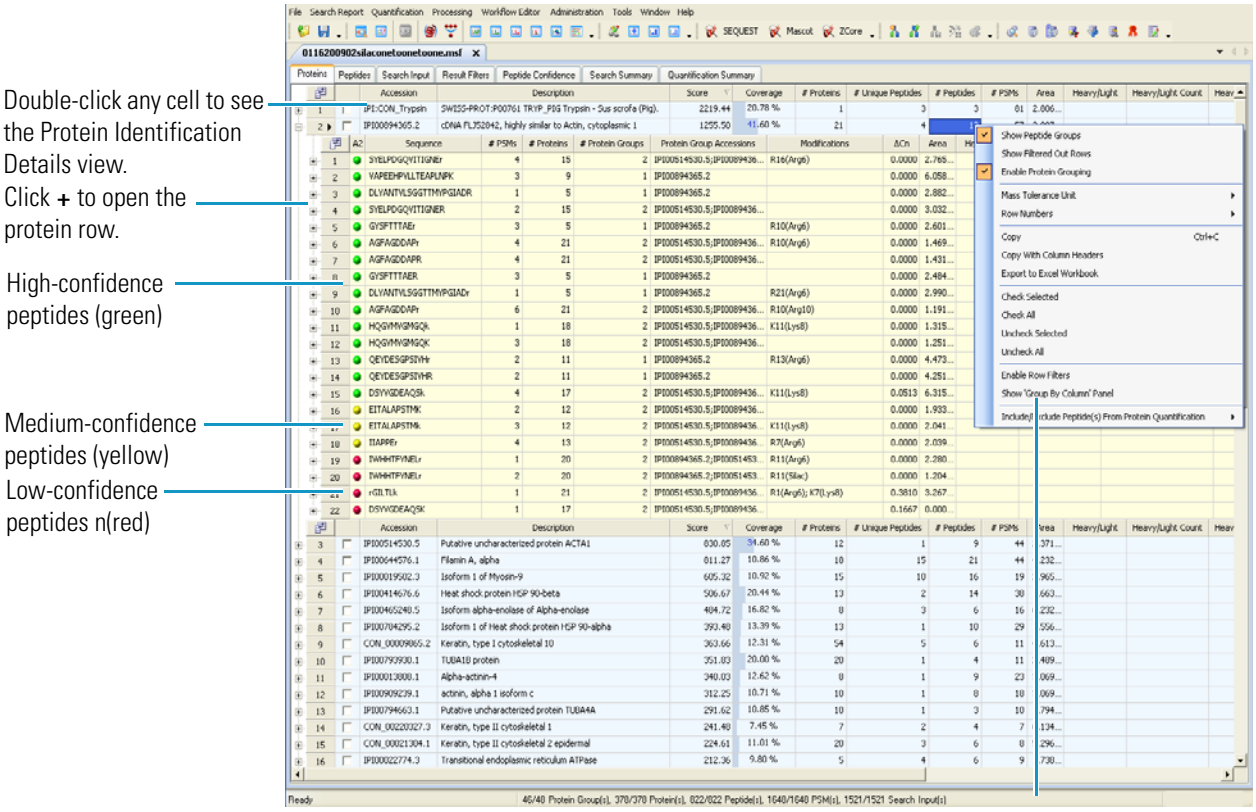
The Proteins page, shown in Figure 144, is displayed by default when you first open the MSF file.

❖ To view the Proteins page

- In an open MSF file, click the **Proteins** page.

The Proteins page of your search report appears, as shown in Figure 144. To select the columns to display, see "Selecting the Columns to Display" on page 223.

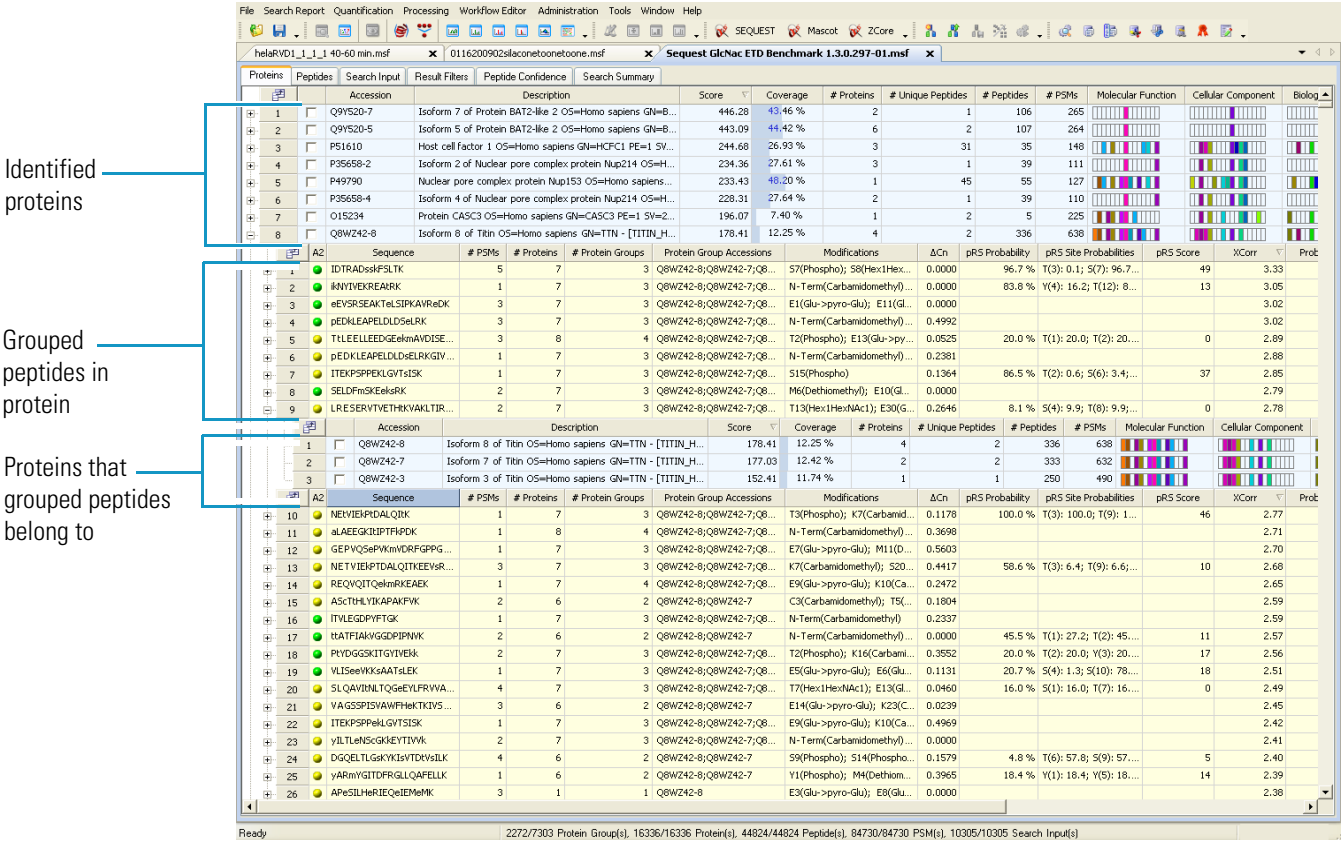
Figure 144. Proteins page



On the Proteins page, you can display three levels when peptides are grouped, as shown in Figure 145:

1. The blue-colored rows on the first level show identified proteins with their associated statistics. Click the plus sign (+) next to a protein row to see the yellow-colored rows.
2. The yellow-colored rows on the second level show the grouped peptides that constitute that protein, as shown in Figure 145. Click the plus sign (+) next to a grouped peptide row to see the blue-colored rows.
   - The columns in this second-level grid show detailed statistics for the peptide associated with the theoretical highest-scoring protein.
   - Each peptide row has a colored circle indicating a confidence level associated with the protein sequence at the top level: green for high confidence, yellow for medium confidence, and red for low confidence.
   - You cannot select peptide group rows, so these rows do not have check boxes.
3. The blue-colored rows on the third level show the proteins that the peptide group belongs to.

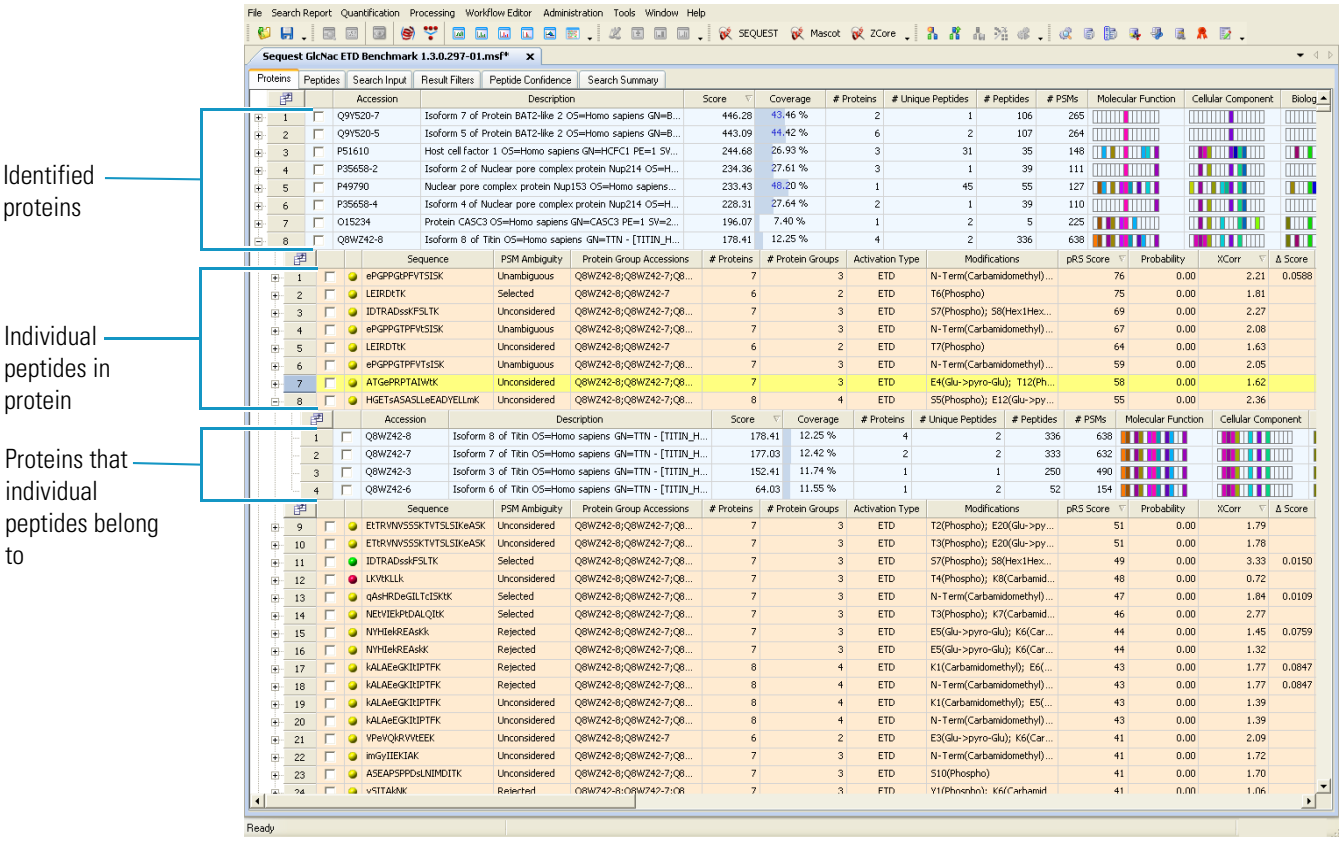
Figure 145. Levels on the Proteins page with grouped peptides



When you right-click a grouped peptide and choose Show Peptide Groups from the shortcut menu to clear the grouping, the peptide rows of the Proteins page appear in orange. For information about the Show Peptide Groups command, see "To group peptides by sequence" on page 228. You can display three levels when peptides are ungrouped, as shown in Figure 146:

1. The blue-colored rows on the first level show identified proteins with their associated statistics. Click the plus sign (**+**) next to a protein row to see the orange-colored rows.
2. The orange-colored rows on the second level show the ungrouped, sorted peptides, as shown in Figure 146. Click the plus sign (**+**) next to a peptide row to see the blue rows.
3. The blue-colored rows on the third level show the proteins that the peptide belongs to.

Figure 146. Levels on the Proteins page with ungrouped peptides



Proteins Page Parameters

Table 39 describes the parameters available on the Proteins page.

Table 39. Proteins page parameters (Sheet 1 of 5)

Feature	Description
Column Chooser ()	Opens the Column Chooser dialog box so you can select which columns to display. An example of this dialog box is shown in Figure 136 on page 223.
Column with check boxes ()	Selects a protein.
Search ID	Displays the order in which the search was submitted to the job queue.
Processing Node No.	Displays the number of the workflow node that was used to identify the peptides composing the protein.
Processing Node Name	Displays the name of the workflow node that was used to identify the peptides composing the protein.
Accession	Displays by default the unique identifier assigned to the protein by the FASTA database used to generate the report.
Description	Provides the name of the protein exclusive of the identifier that appears in the Accession column. This description appears in the table by default.
Score	Displays the protein score, which is the sum of the scores of the individual peptides. • For SEQUEST results, the score is the sum of all peptide Xcorr values above the specified score threshold. The score threshold is calculated as follows: 0.8 + *peptide_charge* × *peptide_relevance_factor* where *peptide_relevance_factor* is an advanced parameter of the SEQUEST node in the "Protein Scoring Option" category with a default value of 0.4. For each spectrum, only the highest-scoring match is used. For each spectrum and sequence, the Proteome Discoverer application uses only the highest scored peptide. When it performs a search using dynamic modifications, one spectrum might have multiple matches because of permutations of the modification site. • For ZCore results, the score is the highest probability score of all peptides identified for that protein.

Table 39. Proteins page parameters (Sheet 2 of 5)

Feature	Description
Score, continued	• For Mascot results, the score is: – Standard score, which is the cumulative protein score based on summing the ion scores of the unique peptides identified for that protein. If a peptide was redundantly identified, only the highest-scoring peptide is used. –or– – MudPIT score, which is the sum of the "excess of ions" score over the homology or identity threshold for each spectrum plus the average threshold. For MudPIT scoring, the score for each peptide is not its absolute score but the amount that it is above the threshold. Therefore, peptides with a score below the threshold do not contribute to the score. For each peptide, the threshold is the homology threshold, if it exists; otherwise, it is the identity threshold. By default, the Proteome Discoverer application automatically switches between the standard and the MudPIT score to calculate the protein score in the Mascot node results. It automatically uses the MudPIT score when the number of queries divided by the number of FASTA database entries exceeds 0.001.
Coverage	Displays by default the percentage of the protein sequence covered by identified peptides.
# Proteins	Displays the number of identified proteins in the protein group of a master protein. This number is the same as that displayed in the Protein Group Members view when you choose Search Report > Show Protein Group Members View (see Figure 156 on page 286).
# Unique Peptides	Displays the number of peptide sequences unique to a protein group.
# Peptides	Displays the number of distinct peptide sequences in the protein group.
# PSMs	Displays the total number of identified peptide sequences (peptide spectrum matches) for the protein, including those redundantly identified.

Table 39. Proteins page parameters (Sheet 3 of 5)

Feature	Description
Molecular Function	Displays the GO Slim categories of the protein's molecular functions as colored boxes, which are explained in Figure 231 on page 440. You must include the Annotation node in the workflow to display this column.
Cellular Component	Displays the GO Slim categories of the protein's cellular components as colored boxes, which are explained in Figure 231 on page 440. You must include the Annotation node in the workflow to display this column.
Biological Process	Displays the GO Slim categories of the protein's biological processes as colored boxes, which are explained in Figure 231 on page 440. You must include the Annotation node in the workflow to display this column.
Pfam IDs	Displays the identification numbers of families of proteins. A special sequence comparison algorithm called the Hidden Markov Model groups proteins into the families by comparing the sequences. Each family has its own ID number that starts with Pf … . For more information on Pfam annotation, see "Pfam Annotation" on page 435.
# AAs	Shows by default the sequence length of the protein.
MW [kDa]	Displays the calculated molecular weight of the protein. The Proteome Discoverer application calculates the molecular weight without considering PTMs. Separating proteins by molecular weight can be one of the steps in two-dimensional gel electrophoresis. You can use the protein's molecular weight as a rough constraint to estimate whether it is reasonable to identify a particular protein in a certain fraction that was searched.

Table 39. Proteins page parameters (Sheet 4 of 5)

Feature	Description
calc. pI	Displays the theoretically calculated isoelectric point, which is the pH at which a particular molecule carries no net electrical charge. The amino acids that make up proteins can be positive, negative, neutral, or polar in nature, and together they give a protein its overall charge. At a pH below their isoelectric point, proteins carry a net positive charge; at a pH above their isoelectric point, they carry a net negative charge. Gel electrophoresis can then separate proteins according to their isoelectric point (overall charge) with a polyacrylamide gel, using a technique called *isoelectric focusing*, which uses a pH gradient to separate proteins. Isoelectric focusing is also the first step in two-dimensional gel polyacrylamide gel electrophoresis. When you have searched the fractions resulting from isoelectric focusing, you can use the calc. pI value to estimate whether you might expect to find a particular protein in the given fraction.
Area (in MSF files with isotopically labeled quantification results only)	Displays the average area of the three unique peptides with the largest peak area.
Heavy/Light (in MSF files with isotopically labeled quantification results only)	Displays the ratio of the quantification values of the heavy and light quantification channels.
Heavy/Light Count (in MSF files with isotopically labeled quantification results only)	Displays the number of peptide ratios that were actually used to calculate a particular protein ratio. For more information about ratio counts, see "Ratio Count" on page 561.
Heavy/Light Variability [%] (in MSF files with isotopically labeled quantification results only)	Displays the variability of the peptide ratios that were used to calculate a particular protein ratio. For more information about protein variability, see "Ratio Variability" on page 561.
Ratio columns (in MSF files with reporter ion quantification results only)	Display the ratio of the intensity of the fragmented tag in a sample to the intensity of the fragmented tag in the control sample.
Ratio Count columns (in MSF files with reporter ion quantification results only)	Displays the number of peptide ratios that were actually used to calculate a particular protein ratio. For more information about ratio counts, see "Ratio Count" on page 561.

Table 39. Proteins page parameters (Sheet 5 of 5)

Feature	Description
Ratio Variability [%] columns (in MSF files with reporter ion quantification results only)	Displays the variability of the peptide ratios that showed how much confidence there is in the calculation of protein ratios. For more information about ratio variability, see "Ratio Variability" on page 561.
Is Master Protein	Indicates whether the protein is the master protein of a protein group. This column appears only in the Protein Group Members view when you choose Search Report > Show Protein Group Members View. For some peptides, a list of proteins might contain this peptide sequence, but none of them is a master protein. This situation can occur if the peptide contains isoleucine at a position where the master protein has leucine, or vice versa, and the proteins are grouped with the Consider Leucine and Isoleucine as Equal option on the Result Filters page.

Clicking the plus (+) sign next to any protein opens the column parameters for the associated peptides. For descriptions of these parameters, see "Peptides Page Parameters" on page 291.

Interpreting Your Results with the Protein Identification Details View

You can use the Protein Identification Details view to quickly gauge how good a match is and to obtain information about the protein.

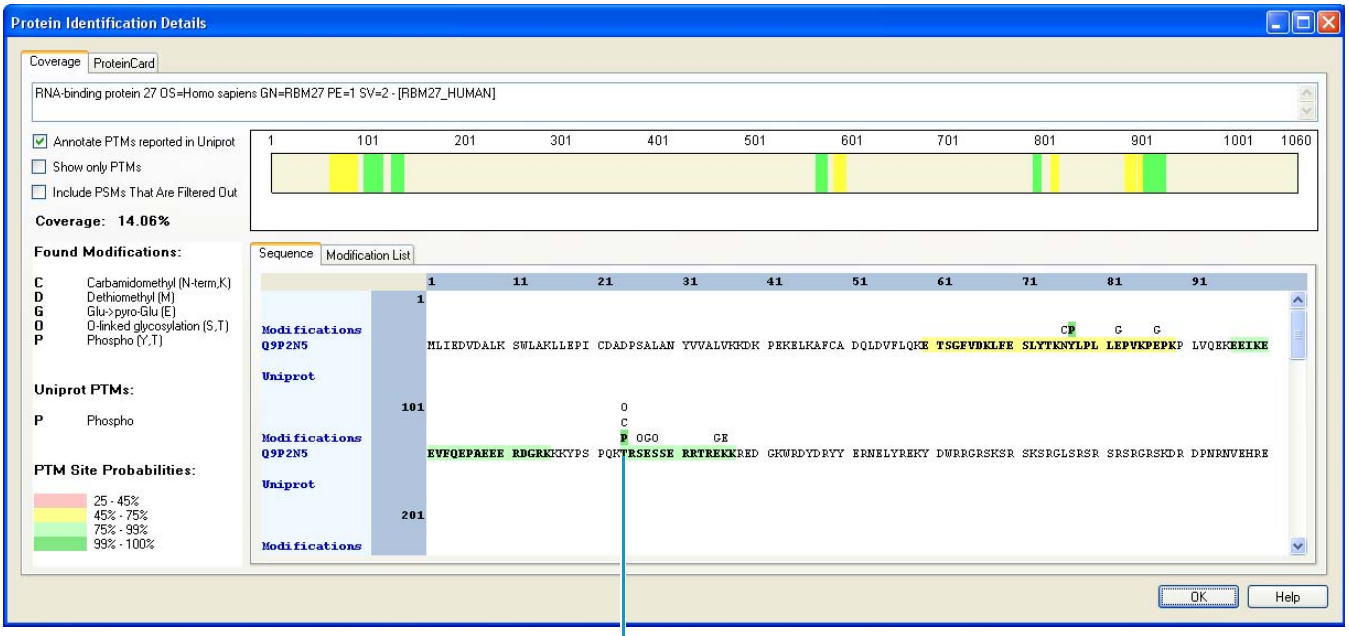
You can also use this view to investigate biologically relevant PTMs. It gives you information about the position of the modification in the peptide and condenses all the different modifications of different identifications of the peptide to the protein sequence. For each protein, the Protein Identification Details view can display information retrieved from ProteinCenter, which accesses publicly available databases such as Uniprot. You can compare the found PTMs to the known PTMs in UniProt to reduce the redundancy of the found PTMs, verify the validity of potential modification sites, and help consolidate the PTM site localization scores that might differ on the level of individual PSMs. (For information on how the phospho*RS* algorithm calculates these scores, see "phosphoRS Node" on page 691.)

The Protein Identification Details view features two pages, Coverage and ProteinCard. For information on accessing the ProteinCard page of the Protein Identification Details view, see "Accessing ProteinCards" on page 449.

Displaying the Coverage Page

The Coverage page, shown in Figure 147, is selected by default.

Figure 147. Multiple modifications on a single amino acid



Multiple modifications on a single amino acid

The left side of the Coverage page displays these features:

- Options that govern the display of sequence annotations of the given protein (for a description of these options, see Table 40)
- Coverage: Displays the percentage of the protein's constituent amino acids that the search found.
- Found Modifications: Lists the modifications found in the search. The Sequence page of the dialog box shows each modification with a single-letter abbreviation and its name and target amino acids. If the Proteome Discoverer application identified more than one modification for an amino acid—for example, a static and a dynamic N-terminal modification—it stacks the abbreviations vertically above the amino acid, as shown in Figure 147 on page 246.
- Uniprot PTMs: Displays the modifications found for the protein in the UniProt database if you used the Annotation or Re-Annotation node in the workflow.
- PTM Site Probabilities: Displays a legend explaining the color-coding of the localization if PTM site localization scores are available (for more information on these scores, see "Sequence Page" on page 247). The site probabilities of the modifications are available if you used the phospho*RS* node in the workflow that generated the results.

The middle section of the Coverage page contains two pages, Sequence and Modification List. The Sequence page is selected by default.

Sequence Page

The Sequence page of the Coverage page of the Protein Identification Details view, shown in Figure 148, includes a protein sequence coverage bar and a sequence table. It displays a graphical and a textual representation of the protein sequence, its coverage with identified peptides, and a list of all found PTMs. The peptide identifications are color-coded, reflecting their respective identification confidence.

Figure 148. Sequence page of the Protein identification Details dialog box



- The protein sequence coverage bar graphically represents the protein sequence. The colored areas represent portions of the protein sequence that have been matched by peptides. The protein sequence coverage bar is colored to indicate the confidence of the peptide sequence identification. The peptides not highlighted were either not detected or not identified.
  - Green indicates high-confidence peptides.
  - Yellow indicates medium-confidence peptides.
  - Red indicates low-confidence peptides.

The numbers on top of the protein sequence coverage bar indicate how many amino acids the protein sequence contains, in increments of 100, and correspond to the numbers in the blue-colored column to the left of the sequence table.

- The sequence table displays the complete amino acid sequence of the currently selected protein. The portions of the sequence that match the searched sequence are highlighted in a color that represents the highest-confidence peptide match in the identification.

  When you click a colored bar in the protein sequence coverage bar, the corresponding sequence is underlined in the sequence table, and an arrow points to the selected bar, as shown in Figure 149.

Figure 149. Colored bar and corresponding underlined sequence

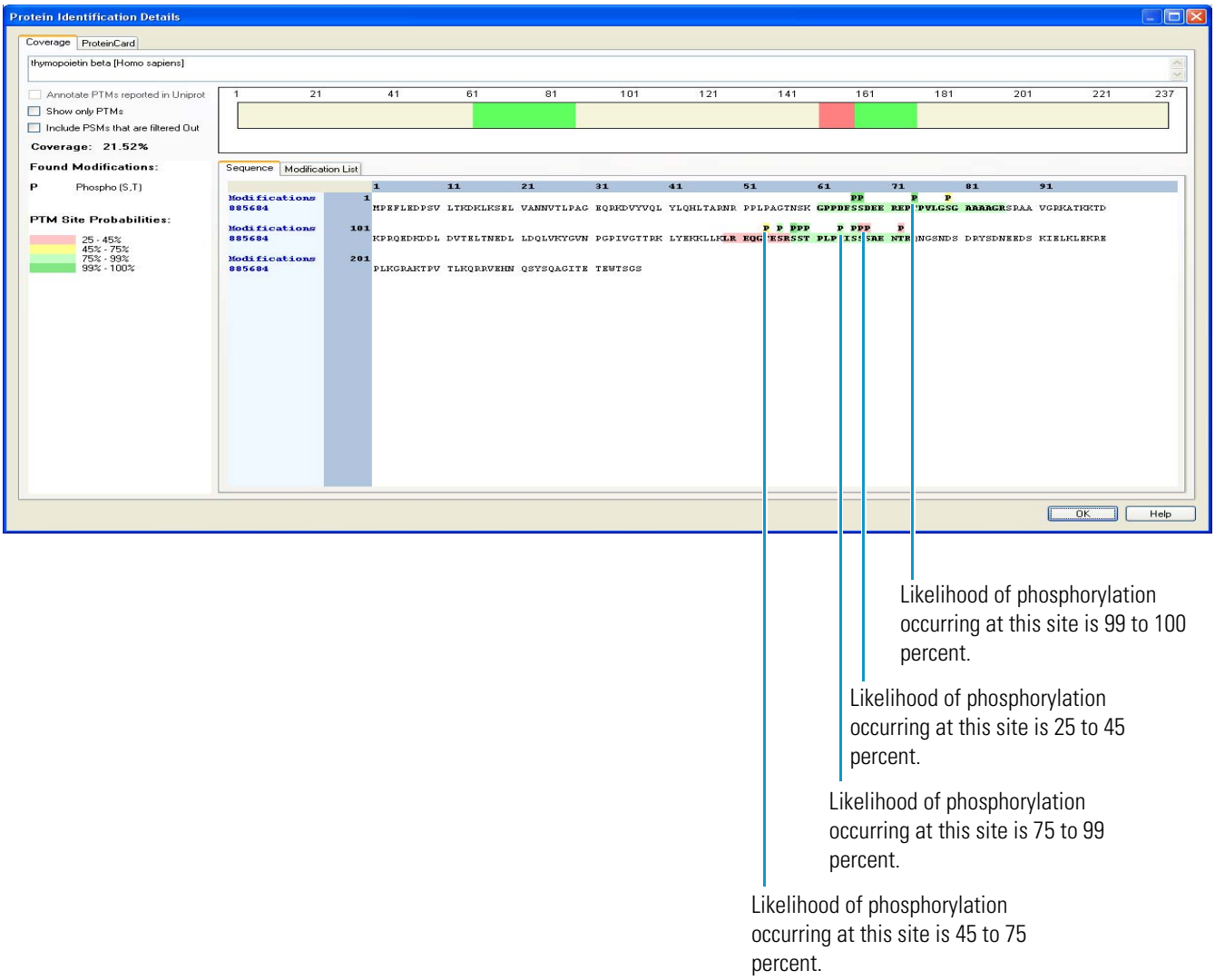


The Proteome Discoverer application annotates the protein sequence with each PTM found in the search by placing the modification's single-letter abbreviation above the sequence, as shown in Figure 148 on page 247. It displays all modifications of PSMs passing the active filters. In some cases, the modification displayed above a PSM of high confidence might only be found in a redundantly identified PSM of medium or low confidence.

When you use the Annotation node (see "Creating a Protein Annotation Workflow" on page 437) or Re-Annotation node (see "Reannotating MSF Files" on page 443) in a workflow, the node retrieves UniProt database information, including PTMs, from ProteinCenter and saves it in Proteome Discoverer results files. If this annotation information is available and you select the Annotate PTMs Reported in UniProt check box, the modification appears as a single-letter abbreviation in gray font in a small gray box beneath the amino acids of the protein in the sequence table, as shown in Figure on page 247.

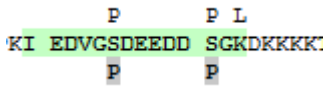
When you use the phospho*RS* node in the workflow (see "Creating a PTM Analysis Workflow" on page 103), PTM site localization scores are available for phosphorylation. In this case, the Proteome Discoverer application displays the found phosphorylation modifications above portions of the amino acid sequences and color-codes the little boxes that they appear in to indicate the probability of the modification being found on those portions. In the PTM Site Probabilities area to the left of the sequence table, it displays the legend explaining the color-coding. Figure 150 shows a phosphorylation color-coded in green and pink boxes to indicate differing probabilities of occurrence at that site.

Figure 150. Differences in color-coded phosphorylations

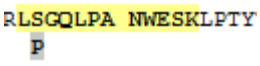


The Proteome Discoverer annotations and the UniProt database annotations together in the sequence table might indicate the following:

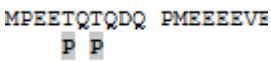
- The modification is documented and found in the MS experiment, which indicates that the protein has been modified.



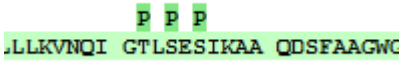
- The modification is documented but not found in the MS experiment, which indicates that the protein might not be modified.



- The modification is documented, but the peptide was not identified in the MS experiment. There are no results, so no information is available about modification. The search engine might have missed the modification because of its settings, so it did not find the peptide.



- The modification is found but not documented in the UniProt database. Either the modification is known but not documented in the database, or the modification at this position is currently unknown and might be worth further investigation.

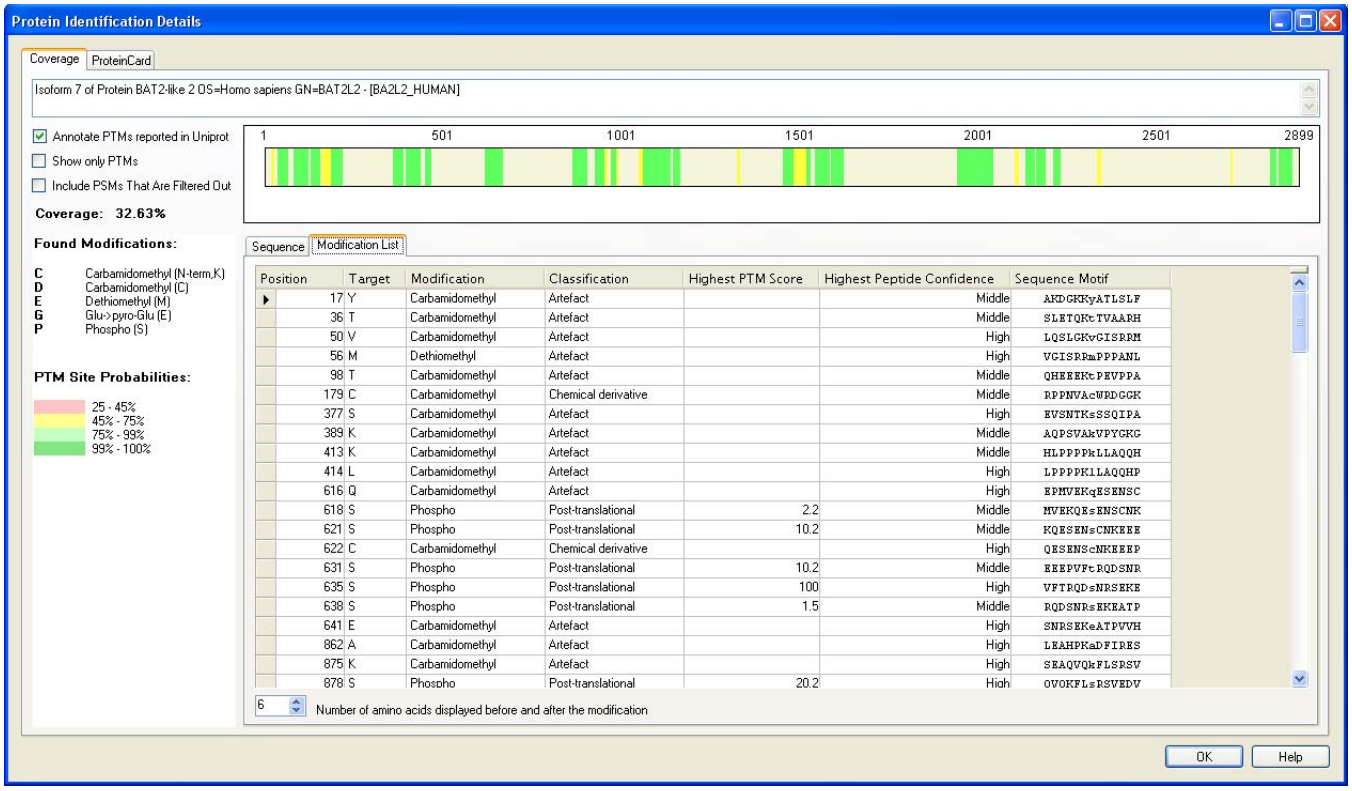


Amino acids in the protein sequence are assigned sequential numbers. The numbers in the column (shown in the blue background in Figure on page 247) to the left of the sequence table and the numbers in the header (shown in the blue background in Figure on page 247) on top of the sequence table are intended to orient you when you want to find automatically assigned amino acid numbers in the table. The numbers in the left column are displayed in increments of 100 and correspond to the numbers on top of the protein sequence coverage bar. The numbers on top of the table are displayed in increments of 10. In the 11-20 column, the first row displays the 11th to 20th amino acid, the second row displays the 111th to 120th amino acid, and so on. For example, the protein sequence shown in Figure 149 contains 698 amino acids, so the numbers to the left of the table extend from 1 to 601, and the numbers across the top of the table extend from 1 to 91.

Modification List Page

The Modification List page of the Coverage Page, shown in Figure 151, lists all the identified post-translational modifications that appear in the sequence table. For information on the columns on this page, see Table 40. You can adjust the length of the sequence motif before and after the modification site by using the Number of Amino Acids Displayed Before and After the Modification option.

Figure 151. Modification List page of the Coverage page of the Protein ID Details view



Using the Protein Identification Details View

The following sections explain how to use the Protein Identification Details view:

- To display a protein sequence
- To display UniProt annotations of the protein
- To show only PTMs
- To display filtered-out PSMs
- To copy the colored bar on the Sequence page
- To save the colored bar on the Sequence page
- To display a protein's post-translational modifications (PTMs)
- To export the Modification List page to Excel

❖ To display a protein sequence

1. Open the MSF file.
2. Double-click a grid cell on the Proteins page, or select a grid cell and choose **Search Report > Show Protein ID Details**, or click the **Show Protein/Peptide ID Details** icon,  .

   The Protein Identification Details view appears.

❖ To display UniProt annotations of the protein

1. Set up and run a workflow that includes the Annotation node or the Re-Annotation node. See "Creating a Protein Annotation Workflow" on page 437 or "Reannotating MSF Files" on page 443.

   Both of these nodes retrieve UniProt as well as GO database information, including PTMs, from ProteinCenter and save them in Proteome Discoverer results files.
2. Open the resulting MSF file, select the protein of interest, and open the Protein Identification Details view.
3. Select the **Annotate PTMs Reported in Uniprot** check box.

   This check box is not available if no annotation information is available because you did not use an Annotation or Re-Annotation node in the workflow. You can find out if the protein has modification information by clicking the ProteinCard tab and then clicking on the Features tab. All lines containing the keyword MOD_RES or CARBOHYD should be displayed as PTMs in the sequence.

   If annotation information is available, the modification appears as a single-letter abbreviation in gray font in a small gray box beneath the amino acids of the protein in the sequence table. Information about the PTMs found in the UniProt database appear under the Uniprot PTMs category in the area to the left of the sequence table. In addition, the likelihood of a PTM occurring on an amino acid appears under the PTM Site Probabilities heading in this same area. Figure 148 on page 247 illustrates this information.

❖ To show only PTMs

- Select the **Show Only PTMs** check box.

  This option displays only PTMs on the Sequence and Modification List pages, so it helps you focus on biologically relevant modifications. It excludes all modifications classified as artifacts, chemical derivatives, isotopic labels, and synthetic peptide protection groups from the display.

❖ To display filtered-out PSMs

- Select the **Include PSMs That Are Filtered Out** check box.

  This option filters out PSMs that have been found but filtered out by the filters on the Results Filters page after the MSF file was opened.

  Note The Include PSMs That Are Filtered Out check box does not display peptides filtered out during the loading of the MSF file.

❖ To copy the colored bar on the Sequence page

- Right-click the bar and choose **Copy**.

❖ To save the colored bar on the Sequence page

1. Right-click the bar and choose **Save As**.
2. In the Save As dialog box, browse to the location where you want to save the file, and type the file name in the File Name box.
3. Click **Save**.

❖ To display a protein's post-translational modifications (PTMs)

1. In the Protein Identification Details view, click the **Coverage** tab.
2. Click the **Modification List** tab.

   The Modification List page displays all identified modifications in the sequence view.

   For information on the columns in this page, see Table 40.
3. In the Number of Amino Acids Before and After Modification box, specify the length of the sequence motif before and after the site of the PTM.

❖ To export the Modification List page to Excel

1. Right-click in the Modification List table and choose **Export to Excel Workbook**.
2. In the Save Modification List as Microsoft Excel File dialog box, browse to the location where you want to save the Excel file, or specify the path and file name in the File Name box.
3. Click **Save**.

Parameters on the Coverage Page of the Protein Identification Details View

Table 40 lists the parameters on the Coverage page of the Protein Identification Details view.

Table 40. Parameters on the Coverage page of the Protein Identification Details view (Sheet 1 of 3)

Command or Option	Description
Description box	Displays the description of the protein. This information is identical to that in the Description column of the Proteins page.
Annotate PTMs Reported in Uniprot	Determines whether to display the PTMs reported in the UniProt database. This option is selected by default. This parameter is not available if no annotation information is available.
Show Only PTMs	Determines whether to display all modifications classified as artifacts, chemical derivatives, isotopic labels, and synthetic peptide protection groups. This option is clear by default.
Include PSMs That Are Filtered Out	Determines whether to display PSMs that were found but filtered out by the filters on the Result Filters page after you opened the MSF file. This option does not display PSMs filtered out during the loading of the MSF file. This option is clear by default.
Coverage	Displays the percent coverage calculated by dividing the number of amino acids in all found peptides by the total number of amino acids in the entire protein sequence. This parameter is read-only.
Found Modifications	Lists all identified PTMs. The list is the same as the list in the Modification List page. Each modification is listed with a single-letter abbreviation, its name, and target amino acids. This parameter is read-only.
Uniprot PTMs	Lists the type of PTMs found in UniProt for the protein.
PTM Site Probabilities	Displays an estimate of the likelihood, in percentages, that phosphorylation will occur on an amino acid, if site localization scores are available from phospho*RS*. For information on how phospho*RS* calculates these scores, see "phosphoRS Node" on page 691.

Table 40. Parameters on the Coverage page of the Protein Identification Details view (Sheet 2 of 3)

Command or Option	Description
Sequence page	
Protein sequence coverage bar	Represents the protein sequence. The colored areas represent portions of the protein sequence that have been matched by peptides. The protein sequence coverage bar is colored to indicate the confidence of the peptide sequence identification. The peptides not highlighted were either not detected or not identified. • Green indicates peptides with high confidence. • Yellow indicates peptides with medium confidence. • Red indicates peptides with low confidence. The numbers on top of the protein sequence coverage bar indicate how many amino acids the protein sequence contains, in increments of 100, and correspond to the numbers in the column to the left of the sequence table.
Sequence table	Displays the complete amino acid sequence of the currently selected protein file, including PTMs. The portions of the sequence that match the searched sequence are highlighted in a color that represents the confidence in the identification. The numbers in the column to the left of the sequence table and the numbers in the header on top of the sequence table are intended to orient you when you want to find automatically assigned amino acid numbers in the table. The numbers in the left column are displayed in increments of 100 and correspond to the numbers on top of the protein sequence coverage bar. The numbers on top of the table are displayed in increments of 10. In the 11-20 column, the first row displays the 11th to 20th amino acid, the second row displays the 111th to 120th amino acid, and so on. For more information on the sequence table, see "Sequence Page" on page 247.
Modification List page	
Position	Displays the position of the amino acid in the protein sequence.
Target	Displays the one-letter code for the modified amino acid.

Table 40. Parameters on the Coverage page of the Protein Identification Details view (Sheet 3 of 3)

Command or Option	Description
Modification	Displays the name of the PTM.
Classification	Displays the classification of the PTM according to Unimod database.
Highest PTM Score	Displays the most likely PTM site as calculated by the algorithm associated with the phospho*RS* node. The Proteome Discoverer application displays this column only if site localization scores are available.
Sequence Motif	Displays the biologically significant amino-acid sequence pattern surrounding the site of the PTM. A lower-case character indicates a modified amino acid.
Highest Peptide Confidence	Displays the highest peptide confidence that the modification is reported in.
Number of Amino Acids Displayed Before and After the Modification	Specifies the number of amino acids that appear before and after the PTM in the Sequence Motif column. The minimum is 0, and the maximum is 100. The default is 6.

Parameters on the ProteinCard Page of the Protein ID Details View

For information on the parameters on the ProteinCard page of the Protein ID details view, see "ProteinCard Parameters" on page 450.

Interpreting Your Results with the Report Item Distribution Chart

You can use the charts in the Report Item Distribution view to plot the properties of the identified peptides and proteins against each other for comparative analysis. You can assess different aspects of the search results, including scores, delta masses, retention times, and so on, by plotting them in relationship to each other. This comparison is available as a scatter plot and as a Venn diagram. In addition, you can use a pie chart to display the relative size of quantities of data.

The Report Item Distribution chart includes an Options pane with several options that you can use to customize each chart. You can save the settings of these options and load them in a future session. You can also return the options in the Options pane to the settings they had when you first installed the Proteome Discoverer application.

The Report Item Distribution chart is interactive and supports multi-level zooming and panning. Use the zoom options to look at the pattern in greater detail.

- Displaying the Report Item Distribution Chart as a Scatter Plot
- Displaying the Report Item Distribution Chart as a Venn Diagram
- Displaying the Report Item Distribution Chart as a Pie Chart

Displaying the Report Item Distribution Chart as a Scatter Plot

A scatter plot shows the distribution of data as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis. A scatter plot is useful when you are trying to refine your methods or to determine whether a mass spectrometer is properly calibrated. You can select or clear all points that are currently visible in the chart, or you can select or clear single points.

- To display a Report Item Distribution chart as a scatter plot
- To return the options to the original default settings
- To zoom in
- To zoom out
- To return to the default view
- To copy the scatter plot
- To save the scatter plot
- To select the visible data points in the scatter plot
- To clear the visible data points in the scatter plot
- To select a single data point in the scatter plot
- To clear a single data point in the scatter plot

❖ To display a Report Item Distribution chart as a scatter plot

1. Choose **Search Report > Show Distribution Chart**, or click the **Show Distribution Chart** icon,  .
2. Click the **Scatter Plots** tab in the Report Item Distribution dialog box.
3. From the Data Source list, select the types of data to display: **Protein Groups**, **All Proteins**, **Peptide Groups**, **PSMs**, or **Search Input**.
4. In the X Data list, select the data set to display on the *x* axis.
5. In the Y Data list, select the data set to display on the *y* axis.
6. (Optional) In the Z Data list, select the data set to display on the *z* axis.

   Selecting a data value for the *z* axis adds a color gradient to the plotted data points ranging from the lowest to the highest Z data value.

7. Move the cursor over the Options label on the left side of the page to temporarily display the Options box. Click in the upper right corner of the Options box to make the box stationary on the page. (Click the **Auto Hide** icon, , to render the Options box temporary again.)

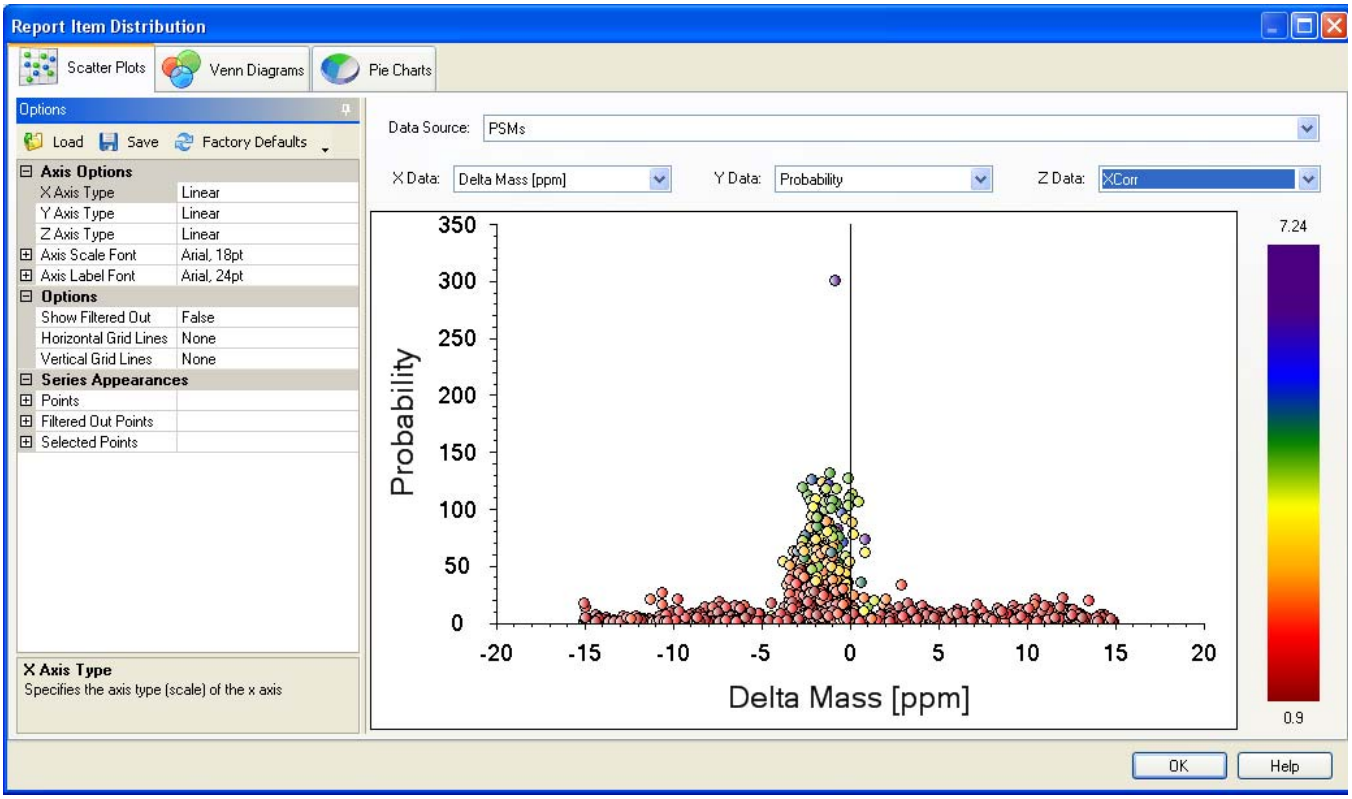
8. Use the options in the Options pane to adjust the colors, labels, and legends of the display. For information on these options, see Table 41.

   –or–

   In the Options pane of the Report Item Distribution chart, click **Load** to load the settings that you most recently saved.

   Figure 152 shows an example of a scatter plot. A scatter plot with this distribution might indicate that you should modify your search workflow to use a ±5 ppm delta mass instead of 15 ppm, because all the best peptides fall within the tighter tolerances.

Figure 152. Scatter Plots page of the Report Item Distribution chart



Moving the cursor over a point in the chart activates a ToolTip with information about the underlying result item.

The default display shows the properties of all peptides currently visible.

ToolTips information that describes the identified peptide, amino acid sequence, charge state, referenced proteins, and so on appears by default. To switch off the display of the ToolTips information, right-click and choose **Show Position Tooltips**.

9. (Optional) To save all the settings that you have chosen, click **Save** in the Options pane.

   The new settings overwrite any previously saved settings. The Proteome Discoverer application loads these new settings when you click Load.

10. To close the Report Item Distribution chart, click **OK**.

❖ To return the options to the original default settings

- In the Options pane of the Report Item Distribution chart, click **Factory Defaults**.

❖ To zoom in

- Drag the cursor to select the area you want to enlarge in size.

❖ To zoom out

- Drag the cursor to the left and select the area you want to reduce in size. You can also right-click and choose **Zoom Out**.

❖ To return to the default view

- Right-click the chart and choose **Undo All Zoom/Pan**.

❖ To copy the scatter plot

- Right-click the chart and choose **Copy**.

  You can paste the contents of the chart into another document.

❖ To save the scatter plot

1. Right-click the chart and choose **Save As**.
2. In the Save As dialog box, browse to the location where you want to save the file, and type the file name in the File Name box.
3. Click **Save**.

❖ To select the visible data points in the scatter plot

1. Zoom in to restrict the view to the area of interest, for example, protein groups with significant up- or down-regulation.
2. Right-click the chart and choose **Select Visible Points.**

   The Proteome Discoverer application places a check mark in the results file for all the result items—protein groups, proteins, peptide groups, PSMs, and search input—corresponding to the selected data points in the chart. However, you must refresh the screen to see the check marks.

   Conversely, you can place a check mark next to the items in the results file that you want to display in a scatter plot.

❖ **To clear the visible data points in the scatter plot**

- Right-click the chart and choose **Deselect Visible Points**.

❖ **To select a single data point in the scatter plot**

- Right-click the data point in the scatter plot, and choose **Select Point**.

❖ **To clear a single data point in the scatter plot**

- Right-click the data point in the scatter plot, and choose **Deselect Point**.

Parameters on the Scatter Plots Page of the Report Item Distribution Chart

Table 41 lists the parameters on the Scatter Plots page of the Report Item Distribution chart.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 1 of 10)

Command or Option	Description
Data Source	Displays the result category to plot the data from: Protein Groups, All Proteins, Peptide Groups, PSMs, or Search Input.
X Data	Specifies the *x*-axis data set. You can select from the following types of data: • # AAs: Sequence length of the protein • # Identified Peptides: Total number of peptide matches found for that spectrum • # Missed Cleavages: Number of cleavage sites in a peptide sequence that a cleavage reagent (enzyme) did not cleave • # Peptides: Number of peptide sequences unique to a protein group • # Protein Groups: Number of protein groups in which this peptide is found • # Proteins: Number of proteins in which this peptide is found • # PSMs: Total number of identified peptide sequences (PSMs) for the protein, including those redundantly identified • % Matched Ions: Percentage of ions matched by the search engine • Calc. pI: Theoretically calculated isoelectric point, which is the pH at which a particular molecule carries no net electrical charge • Charge: Charge state of the peptide • Coverage: Percentage of amino acids found that compose a protein • Δ Cn: Normalized score difference between the currently selected PSM and the highest-scoring PSM for that spectrum • Delta Mass [mmu]: Delta ion masses, in milli-mass units • Delta Mass [ppm]: Delta ion masses, in parts per million

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 2 of 10)

Command or Option	Description
X Data, continued	• Δ Score: Measure of the difference between the top two scores for the peptides identified by that spectrum • First Scan: Number of the first scan used to identify the peptide • Intensity: Intensity of the precursor ion • Ion Inject Time [ms]: Time used to accumulate ions in the mass spectrometer to reach their target value before they are scanned out, in milliseconds • Isolation Interference [%]: Percentage of interference by co-isolation within the precursor isolation window • Last Scan: Number of the last scan used to identify the peptide • Matched Ions: Number of ions matched by the search engine • MH+ [Da]: Protonated monoisotopic mass of the peptides, in daltons • MW [kDa]: Calculated molecular weight of the protein • m/z [Da]: Mass-to-charge ratio of the precursor ion, in daltons • Peptide Length: Sequence of amino acids that compose the peptide • Peptides Matched: Number of peptide candidates that fall into the precursor mass tolerance window set for the search • Precursor Charge: Charge state of the precursor ion • Precursor Intensity: Intensity of the precursor ion • Precursor m/z [Da]: Mass-to-charge ratio of the precursor ion, in daltons • Precursor MH+ [Da]: Singly charged mass of the precursor ions, in daltons

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 3 of 10)

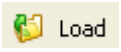
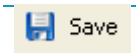
Command or Option	Description
X Data, continued	• Probability: Probability score for the peptide (ZCore and SEQUEST searches only) • pRS Probability: Probability (0–100 percent) that the isoform is correct • pRS Score: Peptide score based on the cumulative binomial probability that the observed match is a random event • Rank: Ordering of peptides by rank • RT [min]: Retention time when the peptide was observed, in minutes • Score: Protein score, which is the sum of the scores of the individual peptides • Search Engine Rank: Original rank assigned by the search engine for all PSMs and peptide groups • SpScore: Raw value of the preliminary score of the SEQUEST algorithm • Total Ions: Total possible number of ions • XCorr: Score of the number of fragment ions that are common to two different peptides with the same precursor mass and the cross-correlation score for all candidate peptides queried from the database (SEQUEST searches only)
Y Data	Specifies the *y*-axis data set. You can select the same types of data as for the *x*-axis data set.
Z Data	Specifies the *z*-axis data set. You can select the same types of data as for the *x*-axis data set.
	Loads the Option pane settings that you saved with the Save command.
	Saves the settings that you selected in the Options pane.
	Resets the settings of the options in the Options pane to the defaults in effect when you installed the Proteome Discoverer application.
Axis Options	
X Axis Type	Specifies the axis type (scale) of the *x* axis. • Linear: Plots the data on a linear scale. • Log: Plots the data on a logarithmic scale.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 4 of 10)

Command or Option	Description
Y Axis Type	Specifies the axis type (scale) of the *y* axis. • Linear: Plots the data on a linear scale. • Log: Plots the data on a logarithmic scale.
Z Axis Type	Specifies the axis type (scale) of the *z* axis. • Linear: Plots the data on a linear scale. • Log: Plots the data on a logarithmic scale.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 5 of 10)

Command or Option	Description
Axis Scale Font	Specifies the font used to label the scale of the axes. The default is Arial, 18 point. Click on the + symbol to the left of Label Font box to display aspects of the label font that you can change: • Name: Specifies the label font. The default font is Arial. • Size: Specifies the size of the label font. The default size is 18. • Unit: Specifies the unit for measuring the size of the label font. The default unit is Point. • Bold: Specifies the weight of the label font. – True: Displays labels in bold font. – (Default) False: Displays labels in normal font. • GdiCharSet: Specifies a byte value that indicates the GDI character set to use with the selected font. The default is 1. • GdiVerticalFont: Gets a Boolean value that indicates whether the selected font is derived from a GDI vertical font. – True: Indicates that the selected font is derived from a GDI vertical font. – (Default) False: Indicates that the selected font is not derived from a GDI vertical font. • Italic: Specifies whether to italicize the selected font. – True: Italicizes the selected font. – (Default) False: Does not italicize the selected font. • Strikeout: Specifies whether to display the selected font shown with strikeout lines through it. – True: Displays the selected font with strikeout lines through it. – (Default) False: Does not display the selected font with strikeout lines through it. • Underline: Specifies whether to underline the selected font – True: Underlines the selected font. – (Default) False: Does not underline the selected font.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 6 of 10)

Command or Option	Description
Axis Label Font	Specifies the font used to label the axes. The default is Arial, 24 point. Click the + symbol to the left of Label Font box to display aspects of the label font that you can change: • Name: Specifies the label font. The default font is Arial. • Size: Specifies the size of the label font. The default size is 24. • Unit: Specifies the unit for measuring the size of the label font. The default unit is Point. • Bold: Specifies the weight of the label font. – True: Displays labels in bold font. – (Default) False: Displays labels in normal font. • GdiCharSet: Specifies a byte value that indicates the GDI character set to use with the selected font. The default is 1. • GdiVerticalFont: Gets a Boolean value that indicates whether the selected font is derived from a GDI vertical font. – True: Indicates that the selected font is derived from a GDI vertical font. – (Default) False: Indicates that the selected font is not derived from a GDI vertical font. • Italic: Specifies whether to italicize the selected font. – True: Italicizes the selected font. – (Default) False: Does not italicize the selected font. • Strikeout: Specifies whether to display the selected font shown with strikeout lines through it. – True: Displays the selected font with strikeout lines through it. – (Default) False: Does not display the selected font with strikeout lines through it. • Underline: Specifies whether to underline the selected font – True: Underlines the selected font. – (Default) False: Does not underline the selected font.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 7 of 10)

Command or Option	Description
Options	
Show Filtered Out	Determines whether to show filtered-out points.
Horizontal Grid Lines	Determines whether the scatter plot displays horizontal grid lines and specifies the style of these lines. • None: Displays no horizontal grid lines on the scatter plot. • Coarse: Displays horizontal grid lines on the scatter plot that are heavier and further apart. • Fine: Displays horizontal grid lines on the scatter plot that are finer and closer together.
Vertical Grid Lines	Determines whether the scatter plot displays vertical grid lines and specifies the style of these lines. • None: Displays no vertical grid lines on the scatter plot. • Coarse: Displays vertical grid lines on the scatter plot that are heavier and further apart. • Fine: Displays vertical grid lines on the scatter plot that are finer and progressively closer together.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 8 of 10)

Command or Option	Description
Series Appearances	
Points	Specifies the appearance of the points in the scatter plot. • Show Points: Determines whether symbols represent the points in the plot. – (Default) True: Symbols represent the points in the plot. – False: Symbols do not represent the points in the plot, so the points are not visible. • Symbol: Specifies the symbol to represent the points in the plot. – (Default) Circle: Displays the points as circles. – Square: Displays the points as squares. – Diamond: Displays the points as diamonds. – Triangle: Displays the points as triangles. – XCross: Displays the points as X-shaped crosses. – Plus: Displays the points as plus signs. – Star: Displays the points as stars. – TriangleDown: Displays the points as inverted triangles. – HDash: Displays the points as horizontal dashes. – VDash: Displays the points as vertical dashes. – UserDefined: Not available in this release. – Default: Not available in this release. – None: Symbols do not represent the points in the plot, so the points are not visible. • Size: Specifies the size of the symbols that represent the points. The default is 8.
Points, continued	• Back Color: Specifies the fill color of the symbols that represent the points. The default is black. • Fore Color: Specifies the outline color of the symbols that represent the points. The default is black.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 9 of 10)

Command or Option	Description
Filtered Out Points	Specifies the appearance of the filtered-out points in the scatter plot. • Show Points: Determines whether symbols represent the filtered-out points in the plot. – (Default) True: Symbols represent the filtered-out points in the plot. – False: Symbols do not represent the filtered-out points in the plot, so the points are not visible. • Symbol: Specifies the symbol to represent the filtered-out points in the plot. – Circle: Displays the points as circles. – Square: Displays the points as squares. – Diamond: Displays the points as diamonds. – Triangle: Displays the points as triangles. – (Default) XCross: Displays the points as X-shaped crosses. – Plus: Displays the points as plus signs. – Star: Displays the points as stars. – TriangleDown: Displays the points as inverted triangles. – HDash: Displays the points as horizontal dashes. – VDash: Displays the points as vertical dashes. – UserDefined: Not available in this release. – Default: Not available in this release. – None: Symbols do not represent the points in the plot, so the points are not visible. • Size: Specifies the size of the symbols that represent the filtered-out points. The default is 8.
Filtered-Out Points, continued	• Back Color: Specifies the fill color of the symbols that represent the filtered-out points. The default is blue. • Fore Color: Specifies the outline color of the symbols that represent the filtered-out points. The default is black.

Table 41. Parameters on the Scatter Plots page of the Report Item Distribution chart (Sheet 10 of

Command or Option	Description
Selected Points	Specifies the appearance of the selected points in the scatter plot. • Show Points: Determines whether symbols represent the selected points in the plot. – (Default) True: Symbols represent the selected points in the plot. – False: Symbols do not represent the selected points in the plot, so the points are not visible. • Symbol: Specifies the symbol to represent the selected points in the plot. – (Default) Circle: Displays the points as circles. – Square: Displays the points as squares. – (Default) Diamond: Displays the points as diamonds. – Triangle: Displays the points as triangles. – XCross: Displays the points as X-shaped crosses. – Plus: Displays the points as plus signs. – Star: Displays the points as stars. – TriangleDown: Displays the points as inverted triangles. – HDash: Displays the points as horizontal dashes. – VDash: Displays the points as vertical dashes. – UserDefined: Not available in this release. – Default: Not available in this release. – None: Symbols do not represent the points in the plot, so the points are not visible. • Size: Specifies the size of the symbols that represent the selected points. The default is 8.
Selected Points, continued	• Back Color: Specifies the fill color of the symbols that represent the selected points. The default is red. • Fore Color: Specifies the outline color of the symbols that represent the selected points. The default is black.

Displaying the Report Item Distribution Chart as a Venn Diagram

A Venn diagram uses circles to represent different sets of data. Each circle represents one data set, and the overlap between the circles represents data items present in each of the data sets. Venn diagrams in the Proteome Discoverer application can display the overlap between two or three data sets. A Venn diagram is useful for comparing approaches. For example, the Venn diagram shown in Figure 153 on page 273 shows the search results found by the Mascot, SEQUEST, and ZCore search engines. It shows how many peptides were found by SEQUEST, how many by Mascot, and how many by ZCore. The circles overlap where more than one search engine found the same peptides. The areas of the circles and overlaps can be of fixed size or proportional to the actual counts. For Venn diagrams, you must have data drawn from at least two comparable variables, such as two mass spectrometers, two fragmentation methods, or two search modes.

- To display a Report Item Distribution chart as a Venn diagram
- To return the options to the original default settings
- To copy a Venn diagram
- To save a Venn diagram

❖ To display a Report Item Distribution chart as a Venn diagram

1. Choose **Search Report > Show Distribution Chart**, or click the **Show Distribution Chart** icon,  .
2. Click the **Venn Diagrams** tab in the Report Item Distribution dialog box.
3. From the Data Source list, select the type of data to display: **Protein Groups**, **All Proteins**, **Peptide Groups**, or **Search Input**. However, not all of these types might be available. The categories of data available in this list depend on the results in the MSF file.
4. From the 1. Group list, select the first data set to compare. From the menu in the Color A box, select the color to display this data set in.
5. From the 2. Group list, select the second data set to compare. From the menu in the Color B box, select the color to display this data set in.
6. (Optional) From the 3. Group list, select the third data set to compare. From the menu in the Color C box, select the color to display this data set in.
7. Move the cursor over the Options label on the left side of the page to temporarily display the Options box. Click  in the upper right corner of the Options box to make the box stationary on the page. (Click the **Auto Hide** icon,  , to render the Options box temporary again.)

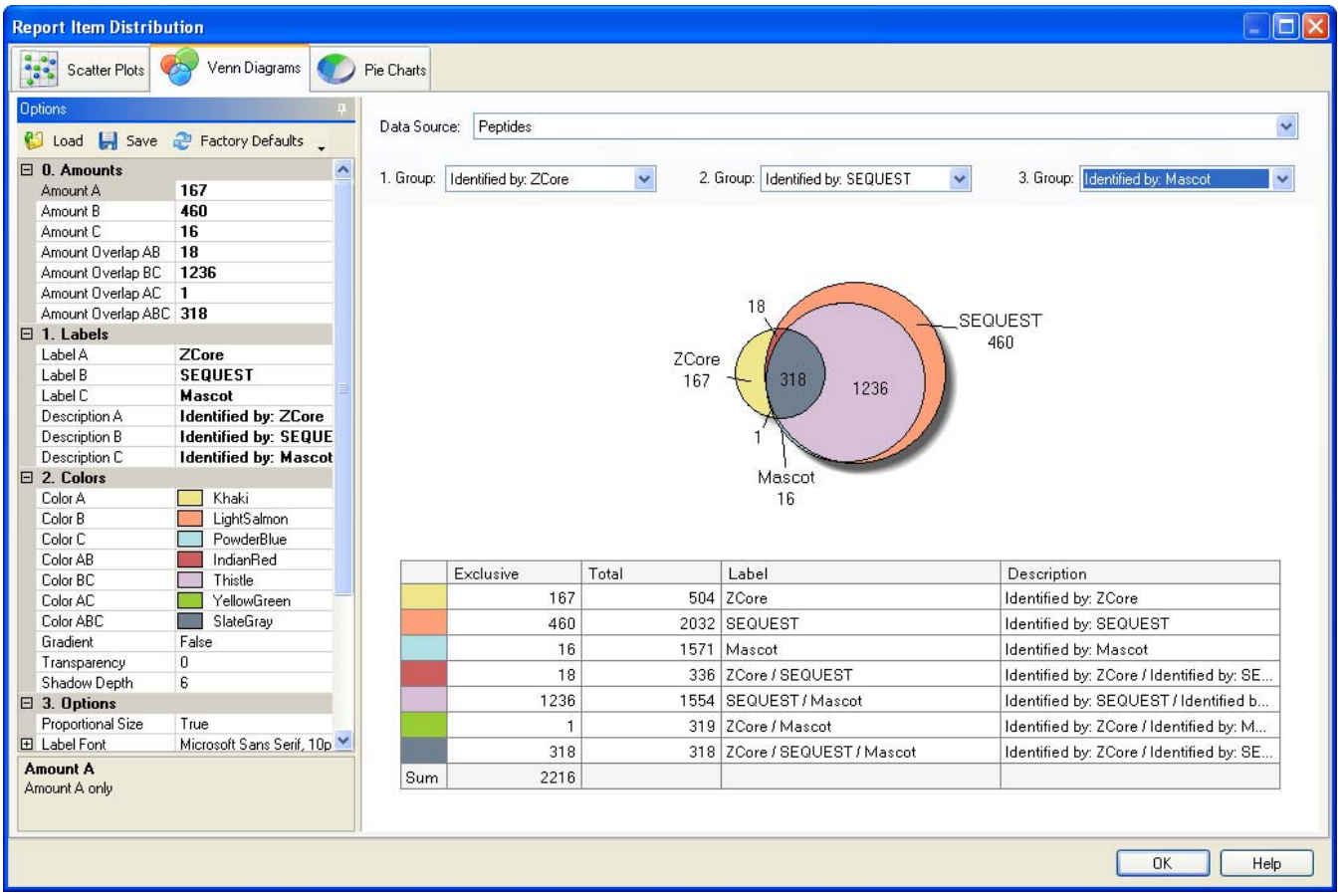
8. Use the options in the Options pane to adjust the colors, labels, and legends of the display. For information on these options, see Table 42.

   –or–

   In the Options pane of the Report Item Distribution chart, click **Load** to load the settings that you most recently saved.

   Figure 153 gives an example of a Venn diagram that compares peptides identified by the ZCore, SEQUEST, and Mascot search engines.

Figure 153. Venn diagram comparing modifications



9. (Optional) To save all the settings that you have chosen, click **Save** in the Options pane.

   The new settings overwrite any previously saved settings. The Proteome Discoverer application loads these new settings when you click Load.

10. To close the Report Item Distribution chart, click **OK**.

❖ To return the options to the original default settings

- In the Options pane of the Report Item Distribution chart, click **Factory Defaults**.

❖ To copy a Venn diagram

- Right-click the chart and choose **Copy**.

  You can paste the contents of the chart into another document.

❖ To save a Venn diagram

1. Right-click the chart and choose **Save**.
2. In the Save As dialog box, browse to the location where you want to save the file, and type the file name in the File Name box.
3. Click **Save**.

Parameters on the Venn Diagrams Page of the Report Item Distribution Chart

Table 42 lists the parameters on the Venn Diagrams page of the Report Item Distribution chart.

Table 42. Parameters on the Venn Diagrams page of the Report Item Distribution chart (Sheet 1 of 5)

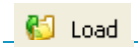
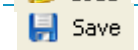
Command or Option	Description
Data Source	Displays the result category to plot the data from: protein groups, proteins, peptide groups, PSMs, or search input.
1. Group	Specifies the first data set to compare.
2. Group	Specifies the second set to compare.
3. Group	Specifies the third data set to compare.
Options box	
	Loads the Option pane settings that you saved with the Save command.
	Saves the settings that you selected in the Options pane.
	Resets the settings of the options in the Options pane to the defaults in effect when you installed the Proteome Discoverer application.
Labels	
Label A	Specifies the label that applies to the data set chosen in 1. Group.
Label B	Specifies the label that applies to the data set chosen in 2. Group.
Label C	Specifies the label that applies to the data set chosen in 3. Group.
Description A	Specifies the legend that applies to the data set chosen in 1. Group. By Default, it displays the name of the data set chosen in 1. Group.
Description B	Specifies the legend that applies to the data set chosen in 2. Group. By Default, it displays the name of the data set chosen in 2. Group.
Description C	Specifies the legend that applies to the data set chosen in 3. Group. By Default, it displays the name of the data set chosen in 3. Group.

Table 42. Parameters on the Venn Diagrams page of the Report Item Distribution chart (Sheet 2 of 5)

Command or Option	Description
Colors	
Color A	Specifies the color for displaying the data set in 1. Group. The default is Khaki.
Color B	Specifies the color for displaying the data set in 2. Group. The default is LightSalmon.
Color C	Specifies the color for displaying the data set in 3. Group. The default is PowderBlue.
Color AB	Specifies the color for displaying the overlap of the 1. Group and 2. Group data sets. The default is IndianRed.
Color BC	Specifies the color for displaying the overlap of the 2. Group and 3. Group data sets. The default is Thistle.
Color AC	Specifies the color for displaying the overlap of the 1. Group and 3. Group data sets. The default is YellowGreen.
Color ABC	Specifies the color for displaying the overlap of the 1. Group, 2. Group, and 3. Group data sets. The default is SlateGray.
Gradient	Determines whether the data sets are displayed in opaque colors or in graduated semi-transparent colors. • True: Displays the data sets in graduated semi-transparent colors. • (Default) False: Displays the data sets in opaque colors.
Transparency	Determines the amount of the fill color in the data set displays. The larger the number that you enter, the more transparent the data set display. The default is 0.
Shadow Depth	Determines the width of the shadow behind the data set displays. The larger the number that you enter, the wider the shadow. The default is 6.
Options	
Proportional Size	Determines whether the area of the data set circles and overlaps are proportional to the number of items in the data set or are of a fixed size. • True: Draws the data set circles and overlaps as proportional to the number of items in the data set. • False: Draws the data set circles and overlaps as fixed sizes.

Table 42. Parameters on the Venn Diagrams page of the Report Item Distribution chart (Sheet 3 of 5)

Command or Option	Description
Label Font	Specifies the font of the data set labels. The default is Microsoft Sans Serif, 10 point. Click on the + symbol to the left of Label Font box to display aspects of the label font that you can change: • Name: Specifies the label font. The default font is Microsoft Sans Serif. • Size: Specifies the size of the label font. The default size is 10. • Unit: Specifies the unit for measuring the size of the label font. The default unit is Point. • Bold: Specifies the weight of the label font. – True: Displays labels in bold font. – (Default) False: Displays labels in normal font. • GdiCharSet: Specifies a byte value that indicates the GDI character set to use with the selected font. The default is 1. • GdiVerticalFont: Gets a Boolean value that indicates whether the selected font is derived from a GDI vertical font. – True: Indicates that the selected font is derived from a GDI vertical font. – (Default) False: Indicates that the selected font is not derived from a GDI vertical font. • Italic: Specifies whether to italicize the selected font. – True: Italicizes the selected font. – (Default) False: Does not italicize the selected font. • Strikeout: Specifies whether to display the selected font shown with strikeout lines through it. – True: Displays the selected font with strikeout lines through it. – (Default) False: Does not display the selected font with strikeout lines through it. • Underline: Specifies whether to underline the selected font – True: Underlines the selected font. – (Default) False: Does not underline the selected font.
Label Color	Specifies the color of the data set labels. The default is black.
Hide Labels	Determines whether the data set labels are visible. • (Default) True: Displays the labels. • False: Hides the labels from view.

Table 42. Parameters on the Venn Diagrams page of the Report Item Distribution chart (Sheet 4 of 5)

Command or Option	Description
Show Legend	Determines whether to display the data set legends. • (Default) True: Displays the data set legends. • False: Does not display the data set legends.
Legend Font	Specifies the font type and size of the legend text. The default is Microsoft Sans Serif, 9 point. Click the + symbol to the left of Legend Font box to display aspects of the label font that you can change: • Size: Specifies the size of the legend font. The default size is 10. • Unit: Specifies the unit for measuring the size of the legend font. The default unit is Point. • Bold: Specifies the weight of the legend font. – True: Displays labels in bold font. – (Default) False: Displays labels in normal font. • GdiCharSet: Specifies a byte value that indicates the GDI character set to use with the selected font. The default is 1. • GdiVerticalFont: Gets a Boolean value that indicates whether the selected font is derived from a GDI vertical font. – True: Indicates that the selected font is derived from a GDI vertical font. – (Default) False: Indicates that the selected font is not derived from a GDI vertical font. • Italic: Determines whether to italicize the selected font. – True: Italicizes the selected font. – (Default) False: Does not italicize the selected font. • Strikeout: Determines whether to display the selected font shown with strikeout lines through it. – True: Displays the selected font with strikeout lines through it. – (Default) False: Does not display the selected font with strikeout lines through it. • Underline: Determines whether to underline the selected font – True: Underlines the selected font. – (Default) False: Does not underline the selected font.
Legend Color	Specifies the color of the legend labels. The default is black.

Table 42. Parameters on the Venn Diagrams page of the Report Item Distribution chart (Sheet 5 of 5)

Command or Option	Description
Border Width	Determines the width of the border around the data set circles. The larger the number that you enter, the wider the line. The default is 1.
Border Color	Specifies the color of the border around the data set circles.

Displaying the Report Item Distribution Chart as a Pie Chart

A pie chart shows several categories of data as slices of a pie (a solid circle composed of slices) or a doughnut (a circle with a hole in the middle) to indicate their relative sizes.

- To display the Report Item Distribution chart as a pie chart
- To return the options to the original default settings
- To copy a pie chart
- To save a pie chart

❖ To display the Report Item Distribution chart as a pie chart

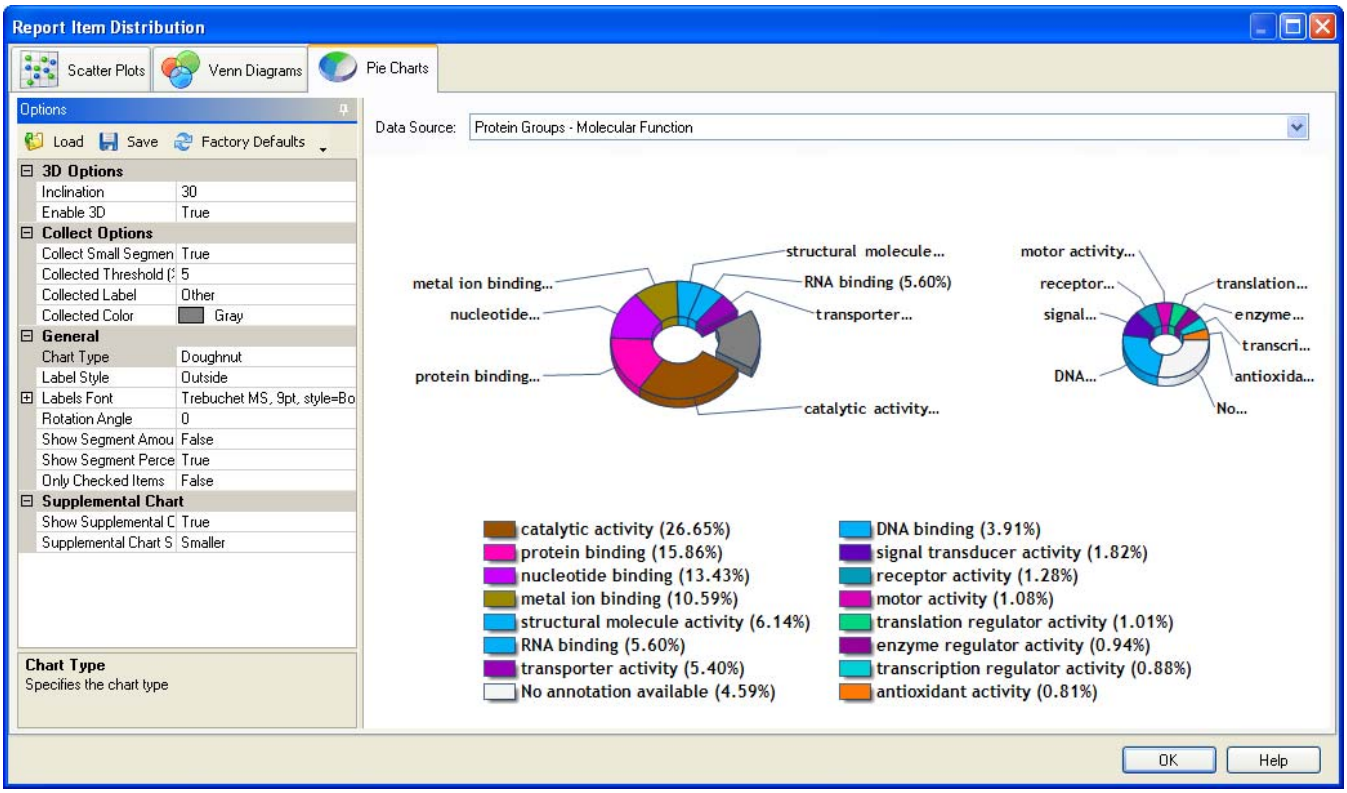
1. Choose **Search Report > Show Distribution Chart**, or click the **Show Distribution Chart** icon,  .
2. Click the **Pie Chart** tab in the Report Item Distribution dialog box.
3. From the Data Source list, select the type of data to display as a pie chart: **Protein Groups**, **All Proteins**, **Peptide Groups**, **PSMs**, or **Search Input**.
4. Move the cursor over the Options label on the left side of the page to temporarily display the Options box. Click  in the upper right corner of the Options box to make the box stationary on the page. (Click the **Auto Hide** icon,  , to render the Options box temporary again.)
5. Use the options in the Options pane to adjust the colors, labels, and legends of the display. Table 43 describes these options.

   –or–

   In the Options pane of the Report Item Distribution chart, click **Load** to load the settings that you most recently saved.

   Figure 154 gives an example of a pie chart that shows the molecular function of each group of proteins in a results file.

Figure 154. Pie chart



6. (Optional) To save all the settings that you have chosen, click **Save** in the Options pane.

   The new settings overwrite any previously saved settings. The Proteome Discoverer application loads these new settings when you click Load.

7. To close the Report Item Distribution chart, click **OK**.

❖ To return the options to the original default settings

- In the Options pane of the Report Item Distribution chart, click **Factory Defaults**.

❖ To copy a pie chart

- Right-click the chart and choose **Copy**.

  You can now paste the contents of the chart into another document.

❖ To save a pie chart

1. Right-click the chart and choose **Save**.
2. In the Save As dialog box, browse to the location where you want to save the file, and type the file name in the File Name box.
3. Click **Save**.

Parameters on the Pie Charts Page of the Report Item Distribution Chart

Table 43 lists the parameters on the Pie Charts page of the Report Item Distribution chart.

Table 43. Parameters on the Pie Charts page of the Report Item Distribution chart (Sheet 1 of 4)

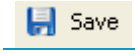
Command or Option	Description
Data Source	Displays the result category to plot the data from: protein groups, proteins, peptide groups, PSMs, or search input.
Options box	
	Loads the Option pane settings that you saved with the Save command.
	Saves the settings that you selected in the Options pane.
	Resets the settings of the options in the Options pane to the defaults in effect when you installed the Proteome Discoverer application.
3D Options	
Inclination	Specifies the angle of inclination, that is, the rotation around the X axis. The default is 30.
Enable 3D	Determines whether the data appears in pseudo-three-dimensional format that uses an inclination. • (Default) True: Displays the data in pseudo-three-dimensional format. • False: Displays the data in a format that uses a color gradient instead of an inclination.
Collect Options	
Collect Small Segments	Determines whether to consolidate small segments together in a single slice. • (Default) True: Consolidates small segments into a single slice. • False: Leaves small segments as is.
Collected Threshold (%)	Specifies the percentage of the total series values. This value determines if the data point value is a “small” value and should be shown as collected. The default is 5 percent.
Collected Label	Specifies the label for the pie segment composed of collected segments. The default label is Other.
Collected Color	Specifies the color of the pie segment composed of collected segments. The default color is Gray.

Table 43. Parameters on the Pie Charts page of the Report Item Distribution chart (Sheet 2 of 4)

Command or Option	Description
General	
Chart Type	Specifies the type of pie chart to display: • Pie: Displays the chart as a solid circle composed of slices. • Doughnut: Displays the chart as a circle with a hole in the middle.
Label Style	Specifies the label style of the chart segments. • Disabled: Does not display a label. • Inside: Displays the label on top of the chart segment. • (Default) Outside: Displays the label outside the chart segment.

Table 43. Parameters on the Pie Charts page of the Report Item Distribution chart (Sheet 3 of 4)

Command or Option	Description
Labels Font	Specifies the font of the data set labels. The default is Microsoft Trebuchet, 9 point bold. Click on the + symbol to the left of Label Font box to display aspects of the label font that you can change: • Name: Specifies the label font. The default font is Microsoft Trebuchet. • Size: Specifies the size of the legend font. The default size is 9. • Unit: Specifies the unit for measuring the size of the legend font. The default unit is Point. • Bold: Specifies the weight of the legend font. – (Default) True: Displays labels in bold font. – False: Displays labels in normal font. • GdiCharSet: Specifies a byte value that indicates the GDI character set to use with the selected font. The default is 1. • GdiVerticalFont: Gets a Boolean value that indicates whether the selected font is derived from a GDI vertical font. – True: Indicates that the selected font is derived from a GDI vertical font. – (Default) False: Indicates that the selected font is not derived from a GDI vertical font. • Italic: Determines whether to italicize the selected font. – True: Italicizes the selected font. – (Default) False: Does not italicize the selected font. • Strikeout: Determines whether to display the selected font shown with strikeout lines through it. – True: Displays the selected font with strikeout lines through it. – (Default) False: Does not display the selected font with strikeout lines through it.
Rotation Angle	Specifies the location of the first chart slice.

Table 43. Parameters on the Pie Charts page of the Report Item Distribution chart (Sheet 4 of 4)

Command or Option	Description
Supplemental Chart	
Show Supplemental Chart	Determines whether to display small segments in a supplemental pie chart. • (Default) True: Displays small segments in a supplemental pie chart. • False: Leaves small segments as is.
Supplemental Chart Size	Specifies the size of the supplemental pie chart relative to the original pie chart. • Largest: Displays the supplemental pie chart as much larger than the original pie chart. • Larger: Displays the supplemental pie chart as larger than the original pie chart. • Comparable: Displays the supplemental pie chart as about the same size as the original pie chart. • (Default) Smaller: Displays the supplemental pie chart as smaller than the original pie chart. • Smallest: Displays the supplemental pie chart as much smaller than the original pie chart.

Interpreting Your Results with the Sequence Comparison View

You can display the complete sequences for different proteins with the Sequence Comparison view. You might want to use this feature to compare similar proteins that share varying degrees of identical peptide sequences or to compare the complete sequences of different proteins associated with a single peptide.

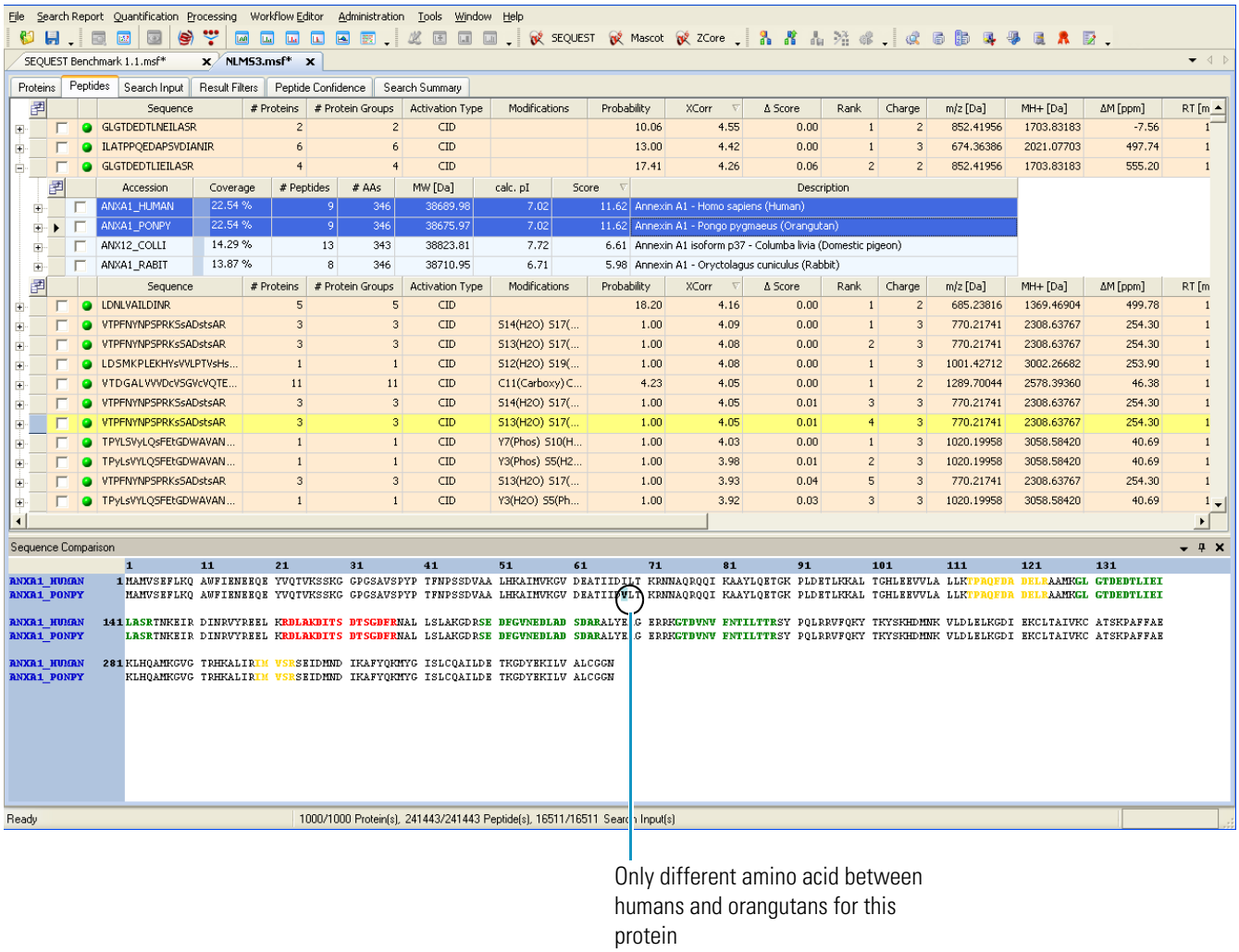
❖ To display the Sequence Comparison view

1. In an open report, click the **Proteins** page, or click the **Peptides** page and then click + to the left of a peptide to show one or more protein rows.
2. Right-click anywhere on the page and turn off **Enable Protein Grouping**, if it is not already turned off.
3. Look in the #Proteins column to find peptides that have more than one protein.
4. Highlight the protein rows of the proteins that you want to compare. For example, in Figure 155, the Annexin A1 proteins of the human and the orangutan are highlighted for comparison.

5. Choose **Search Report > Show Sequence Comparison**, or click the **Show Sequence Comparison** icon, .

   The Sequence Comparison view appears, as shown in Figure 155.

Figure 155. Sequence Comparison view



The sequence table shows the complete amino acid sequence of the currently selected protein files. The portions of the sequence that match the searched sequence are highlighted in a color that represents the confidence in the identification:

- Green: Highlighted peptides of high confidence
- Yellow: Highlighted peptides of medium confidence
- Red: Highlighted peptides of low confidence

The blue column to the left of the table displays the accession numbers of the highlighted protein sequences being compared. In the example in Figure 155, the accession numbers are ANXA1_HUMAN and ANXA1_PONPY. The numbers in the blue-colored column to the left of the sequence table and the numbers in the blue-colored header on top of the sequence table help you find automatically assigned amino acid numbers in the table. The numbers in the left column are displayed in increments of 100, and the numbers on top of the table are displayed in increments of 10.

In the example in Figure 155, the sequence table shows that only one amino acid, V (number 68), is different between the two proteins for humans and orangutans. It is highlighted in blue.

Sequence Comparison View Parameters

Table 44 lists the parameters on the Sequence Comparison view.

Table 44. Sequence Comparison view parameters

Parameter	Description
Sequence table	Displays the complete amino acid sequence of the currently selected protein files. The portions of the sequence that match the searched sequence are highlighted in a color that represents the level of confidence in the identification. • Green: Highlighted peptides of high confidence • Yellow: Highlighted peptides of medium confidence • Red: Highlighted peptides of low confidence The numbers in the blue-colored column to the left of the sequence table and the numbers in the blue-colored header on top of the sequence table help you find automatically assigned amino acid numbers in the table. The numbers in the left column are displayed in increments of 100, and the numbers on top of the table are displayed in increments of 10.
Blue column to the left of the sequence table	Displays the accession numbers of the highlighted protein sequences being compared.

Interpreting Your Results with the Protein Group Members View

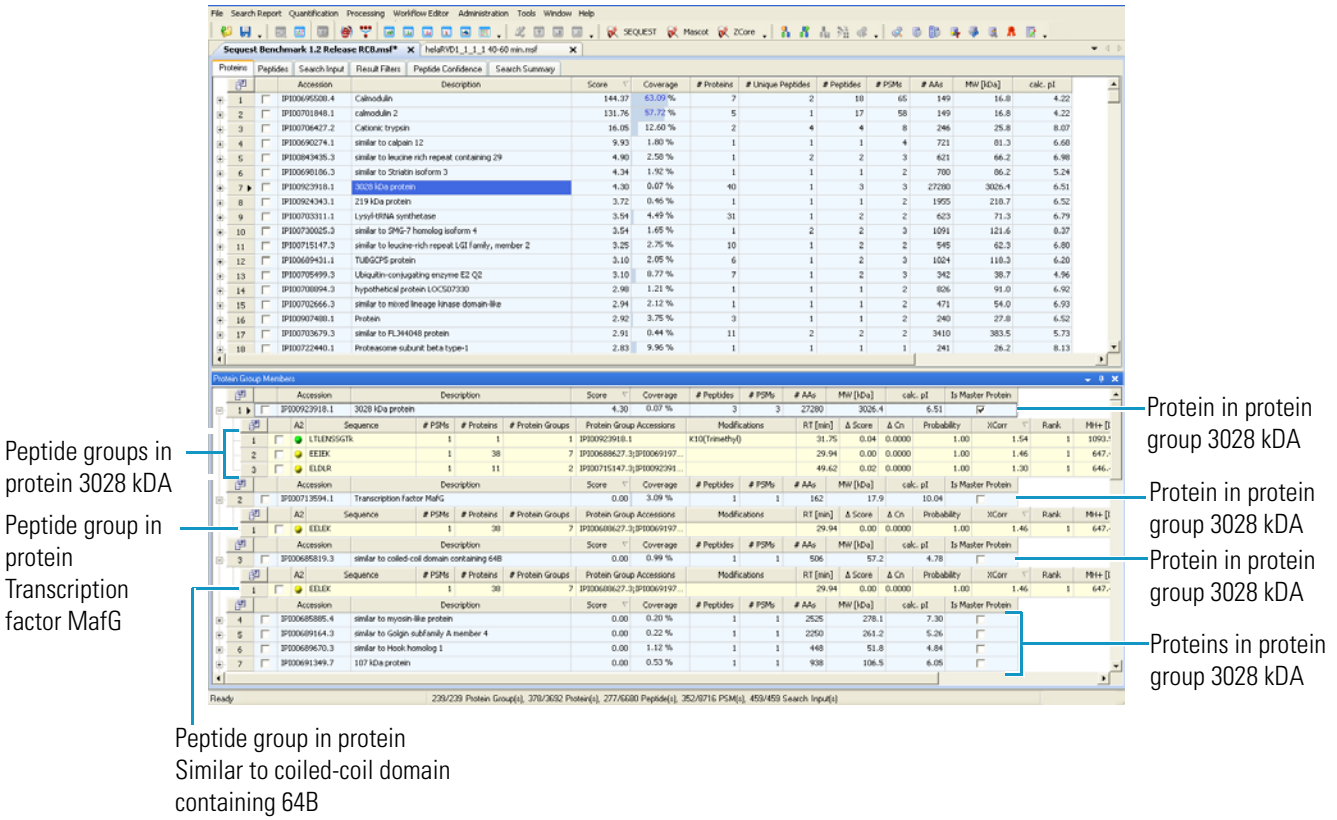
When you enable protein grouping, the Proteome Discoverer application groups proteins that include any of the same peptides. You can use the Protein Group Members view to show which proteins are included in a protein group. For more information on protein grouping, see "Grouping Proteins" on page 401.

❖ To display the Protein Group Members view

1. Right-click the Proteins page of the MSF file and choose **Enable Protein Grouping**, if it is not already selected.
2. Click a protein row.
3. Choose **Search Report > Show Protein Group Members View** or click the **Show Protein Group Members View** icon, .

   The Protein Group Members view opens at the bottom of the Proteome Discoverer window, as shown in Figure 156. In this example, the view shows the proteins in the 3028 kDa protein group. It also shows the grouped peptides that belong to these proteins.

Figure 156. Protein Group Members view



For information on the columns of the Protein Group Members view, including the Is Master Protein column, see "Proteins Page Parameters" on page 241.

When protein grouping is not enabled (that is, when the Enable Protein Grouping command is not selected) the Protein Group Members view displays just the protein selected.

Interpreting Your Results with the Protein References of a Peptide View

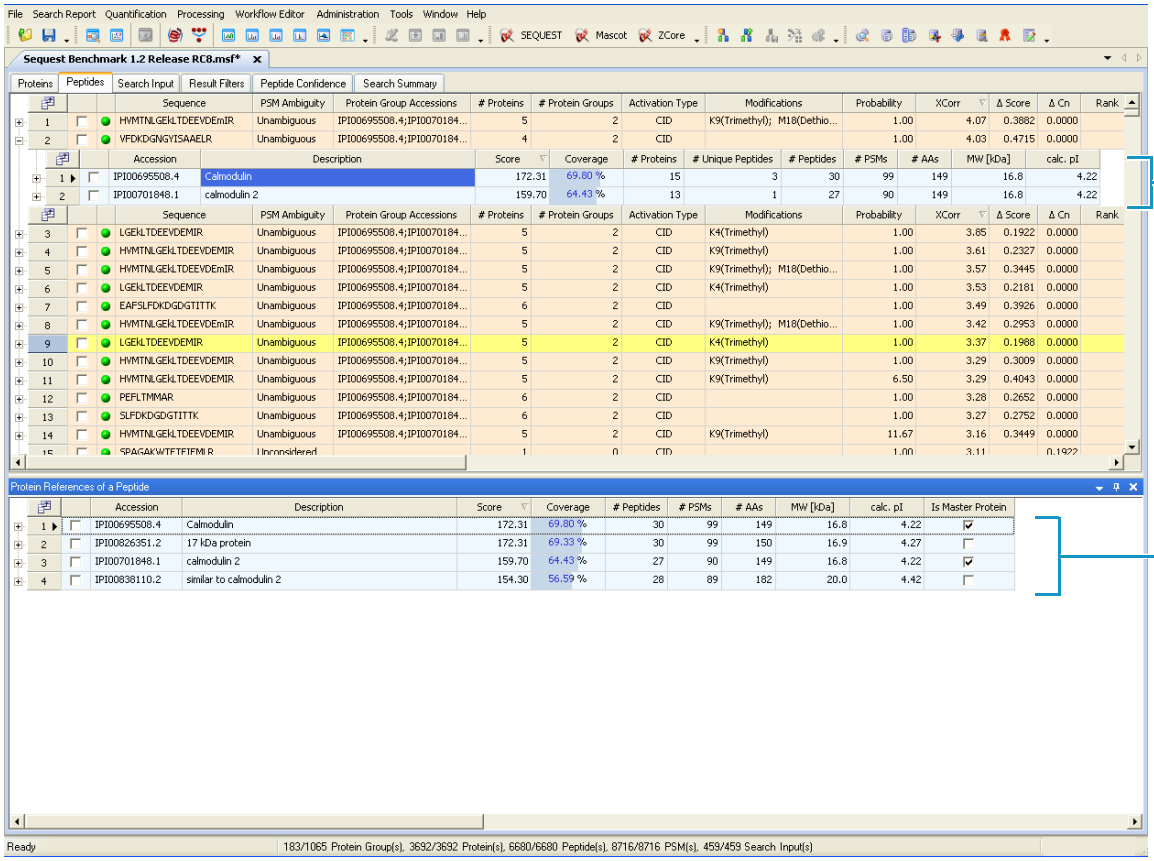
The Protein References of a Peptide view displays the proteins that a specific peptide belongs to.

❖ To display the Protein References of a Peptide view

1. Click in a peptide row on the Peptides or Proteins page.
2. Choose **Search Report > Show Protein References**.

   The Protein References of a Peptide view opens at the bottom of the Proteome Discoverer window, as shown in Figure 157. In this example, the view shows the individual proteins that the VFDKDGNGYISAAELR peptide belongs to. In contrast, the blue-colored rows below the peptide display the protein groups that the peptide belongs to.

Figure 157. Protein References of a Peptide view



For information on the columns of the Protein References of a Peptide view, including the Is Master Protein column, see "Proteins Page Parameters" on page 241.

Using the Peptides Page

The Peptides page, shown in Figure 158 on page 289, lists all the peptides and associated proteins found in a sample. By default, it displays grouped peptides, which appear in yellow rows. All displayed peptides are above the threshold set in any applied filters. You can display peptides below the threshold in unavailable form if you select the Show Filtered Out Rows command from the shortcut menu. For more information about displaying peptides, see "Displaying Peptides" on page 224.

Use the Peptides page to perform most of your tabular and graphical analyses. You can also access various detailed views for peptide matches.

- Displaying the Peptides Page
- Interpreting Your Results with the Peptide Identification Details View
- Interpreting Your Results with the Report Item Distribution Chart
- Interpreting Your Results with the Peptide Consensus View
- Interpreting Your Results with the Dual-Peptide Consensus View
- Displaying Post-Translational Modification Information in the UNIMOD Web View
- Interpreting Your Results with the Chromatogram View
- Interpreting Your Results with the Spectrum View
- Interpreting Your Results with the Fragment Match Spectrum View
- Interpreting Your Results with the Precursor Isotope Pattern View
- Interpreting Your Results with the Extracted Ion Chromatogram View
- Interpreting Your Results with the Peptide Group Members ViewCalculating the Delta Score

Displaying the Peptides Page

You can access the Peptides page from an open results report.

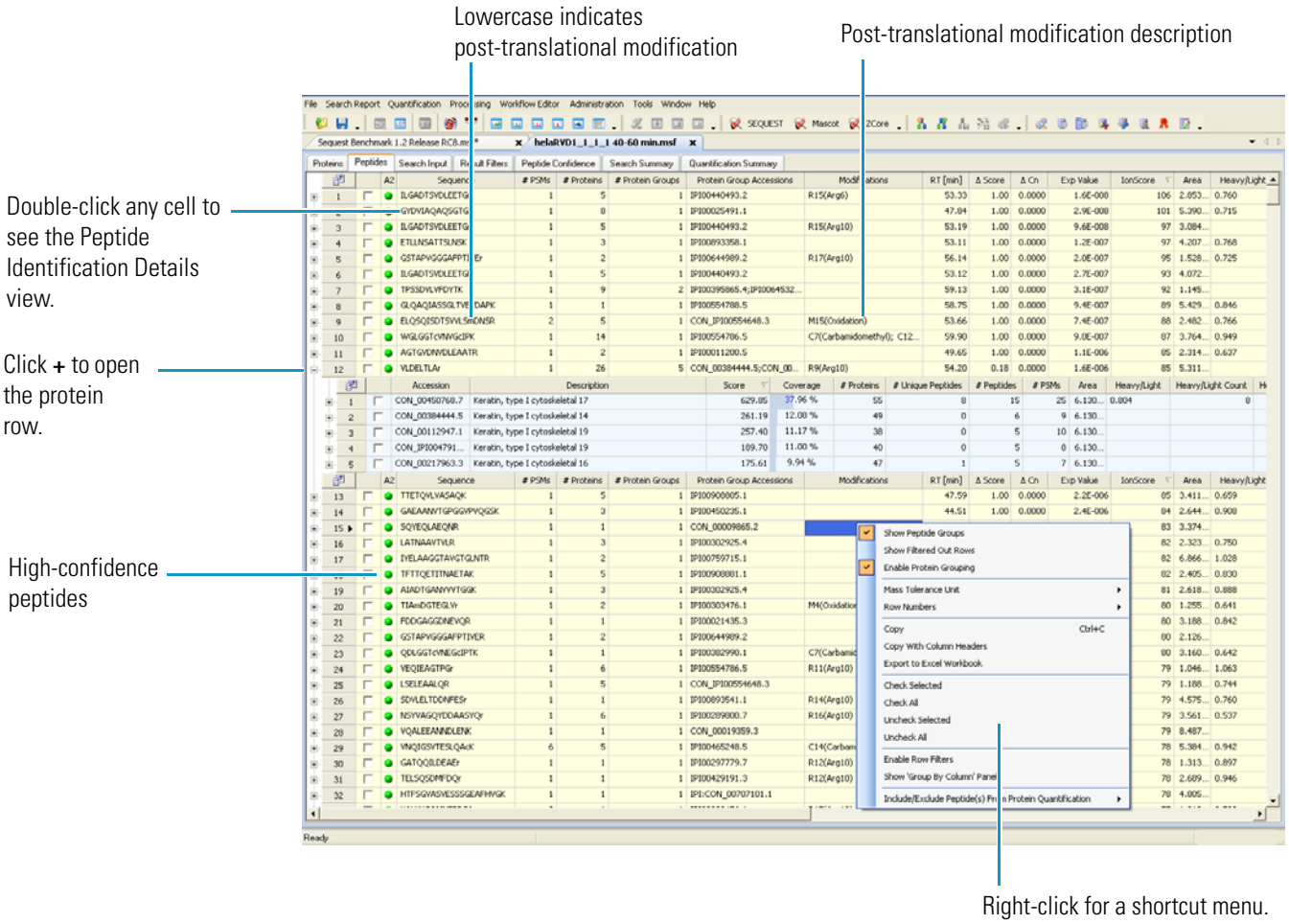
❖ To display the Peptides page

- In an open MSF file, click the **Peptides** tab.

  The Peptides page of your search report appears, as shown in Figure 158, listing all the peptide groups found in the sample. To select the columns to display, see "Selecting the Columns to Display" on page 223.

  If the peptides are not already grouped, you can group them by right-clicking the Peptides page and choosing Show Peptide Groups.

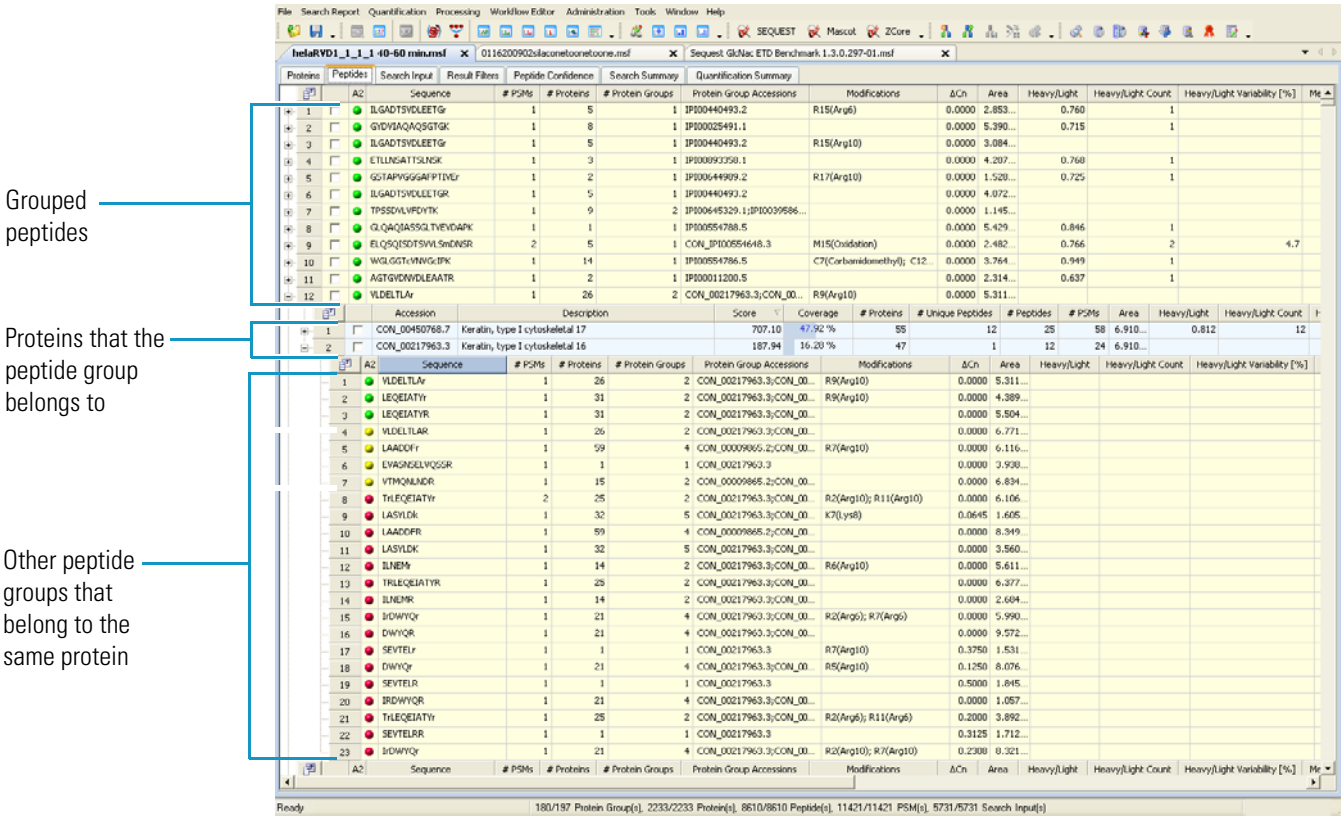
Figure 158. Peptides page



On the Peptides page, you can display three levels, as shown in Figure 159:

1. The yellow-colored rows on the first level show the grouped peptides. Click the plus sign (+) to the left of a peptide group on this level to display the blue rows.
2. The blue-colored rows on the second level show the proteins that a particular peptide group belongs to. Click the plus sign (+) to the left of a protein on this level to display the yellow rows.
3. The yellow-colored rows on the third level show the other peptides that belong to this same protein. You cannot select peptide group rows, so these rows do not have check boxes.

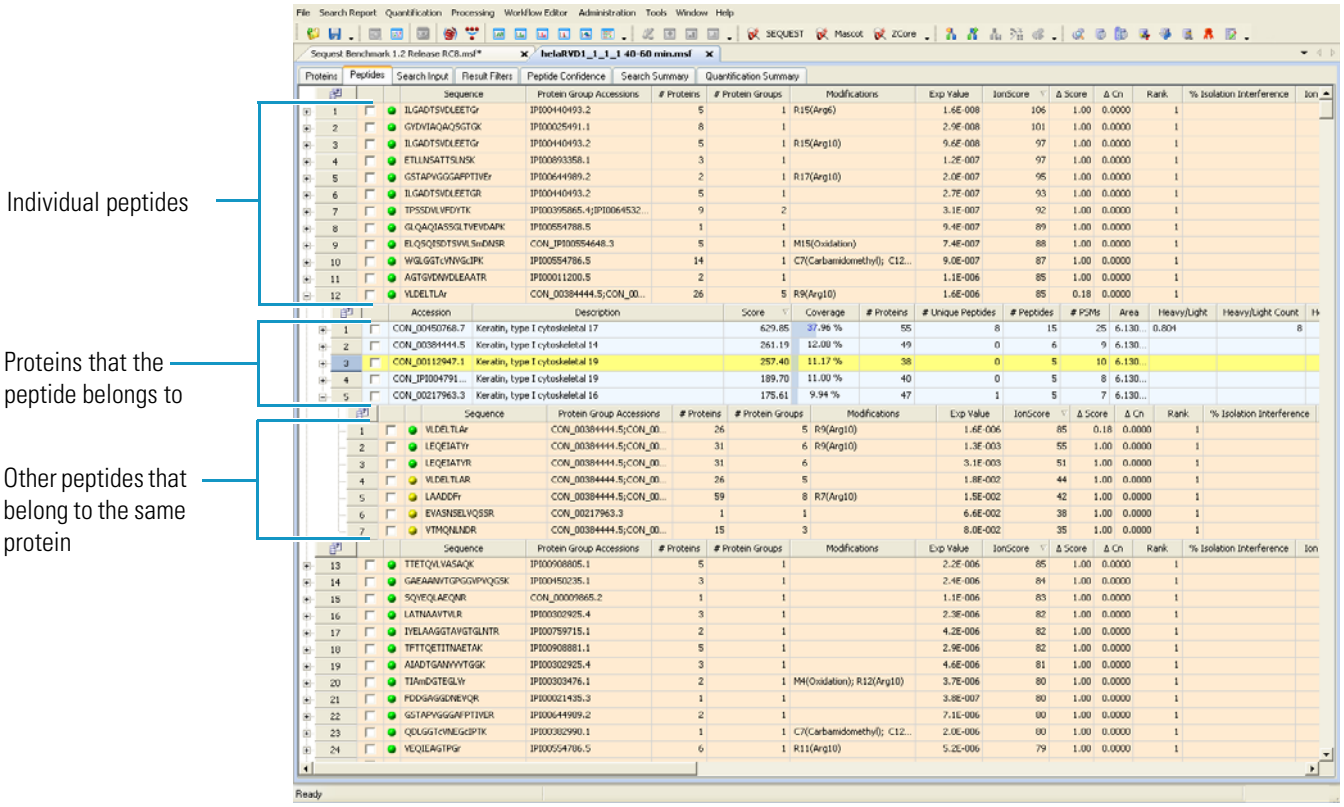
Figure 159. Levels on the Peptides page with grouped peptides



When you right-click and choose Show Peptide Groups from the shortcut menu to deselect peptide grouping, the peptide rows of the Peptides page appear in orange, as shown in Figure 160. For information about the Show Peptide Groups command, see "To group peptides by sequence" on page 228. You can display three levels when peptides are ungrouped, as shown in Figure 146:

1. The orange-colored rows on the first level show the individual peptides. Click the plus sign (+) to the left of a peptide group on this level to display the blue rows.
2. The blue-colored rows on the second level show the proteins that a particular peptide belongs to. Click the plus sign (+) to the left of a protein on this level to display the yellow rows.
3. The orange-colored rows on the third level show the other peptides that belong to this same protein.

Figure 160. Levels on the Peptides page with ungrouped peptides



Peptides Page Parameters

Table 45 describes the parameters on the Peptides page.

Table 45. Peptides page parameters (Sheet 1 of 9)

Feature	Description
Column Chooser ( )	Opens the Column Chooser dialog box so you can choose which columns to display.
Column with check boxes ( )	Allows you to select a peptide.
Confidence icon	Indicates a confidence level associated with the peptide sequence at the top level: green for high confidence, yellow for medium confidence, and red for low confidence.

Table 45. Peptides page parameters (Sheet 2 of 9)

Feature	Description
PSM Ambiguity	Displays the grouping status of the PSM. This column can display the following categories: • Unambiguous: Indicates that this PSM is the only match that the Proteome Discoverer application considered for this spectrum. There is no ambiguity that it needs to resolve. • Selected: Indicates that the Proteome Discoverer application selected this PSM from a set of two or more matches that it considered for the protein group inference process. • Rejected: Indicates that the Proteome Discoverer application rejected this PSM from a set of two or more matches that it considered for the protein group inference process. • Ambiguous: Indicates that the Proteome Discoverer application considered two or more peptide matches to the same spectrum for the protein group inference process. If you select the Apply Strict Maximum Parsimony Principle option in the Protein Grouping (Enabled) area of the Result Filters page, the application resolves this ambiguity to Selected or Rejected. • Unconsidered: Indicates that the Proteome Discoverer application did not consider this PSM for the protein group inference process because it did not meet the requirements specified in the Protein Grouping (Enabled) settings in the Result Filters page. For more information on the protein group inference process that the application uses to assign the grouping status of PSMs, see "Grouping Proteins" on page 401.
Search ID	Displays the order in which the search was submitted to the job queue.
Processing Node No.	Displays the number of the workflow node that was used to identify the peptides composing the protein.
Sequence	Displays the sequence of amino acids that compose the peptide.

Table 45. Peptides page parameters (Sheet 3 of 9)

Feature	Description
# PSMs	Displays the total number of identified peptide sequences (PSMs) for the protein, including those redundantly identified.
Protein Group Accessions	Displays the unique identifiers (accessions) of all master proteins from all protein groups that include this peptide sequence. If you group proteins by selecting the Consider Leucine and Isoleucine as Equal option on the Result Filters page when you click on Settings under Protein Group (Enabled), this column also lists identifiers from master proteins that in many cases include this specific peptide sequence. The identifiers displayed in the Protein Group Accessions column are the same as those displayed in the Accession column on the Proteins page. They are also the same identifiers displayed when you click + to the left of a peptide on the Peptides page to display proteins.
Protein Descriptions	Describes the proteins associated with the peptide. This description is taken from the FASTA file.
# Proteins	Displays the number of proteins in which this peptide is found.
# Protein Groups	Displays the number of protein groups in which this peptide is found.
Activation Type	Displays the activation type of the spectrum in which the peptide was identified.
Modifications	Displays the static and dynamic modifications identified in the peptide.
Probability	Displays the probability score for the peptide (ZCore and SEQUEST searches only).
SpScore (search-dependent)	Displays the raw value of the preliminary score of the SEQUEST algorithm (SEQUEST searches only).
XCorr (search-dependent)	Scores the number of fragment ions that are common to two different peptides with the same precursor mass and calculates the cross-correlation score for all candidate peptides queried from the database (SEQUEST searches only).

Table 45. Peptides page parameters (Sheet 4 of 9)

Feature	Description
Expectation Value or Exp Value (Mascot and ZCore only)	Displays the expectation value, which is a measure of the number of matches with scores equal to or better than the score values that are expected to occur only by chance. For Mascot, the smaller the expectation value is, the better the match is considered.
IonScore (Mascot only)	Refer to the Mascot documentation from Matrix Science.
Δ Score	Displays a measure of the difference between the top two scores for the peptides identified by that spectrum. The Proteome Discoverer application calculates this score as ΔScore = Score(Rank *N* Peptide) – Score(Rank 1 Peptide)/Score(Rank 1 Peptide). For a detailed discussion of the delta score calculation, see "Calculating the Delta Score" on page 345.
Δ Cn	Displays the normalized score difference between the currently selected PSM and the highest-scoring PSM for that spectrum. For more information on how the Proteome Discoverer application calculates the Delta Cn value, see "Filtering Peptides by Delta Cn Value" on page 385.
Rank	Displays the ordering of peptides by rank.
Search Engine Rank	Displays the original rank assigned by the search engine for all PSMs and peptide groups. For more information on this parameter, see "Filtering Results by the Original Rank Assigned by the Search Engine" on page 387.
# Missed Cleavages	Displays the number of cleavage sites in a peptide sequence that a cleavage reagent (enzyme) did not cleave. This number excludes cases where an amino acid such as proline inhibits the cleaving enzyme. For example, if proline resides next to lysine or arginine, trypsin does not cleave the lysine or arginine. The # of Missed Cleavages value is 0 for a complete digest of all peptides.
% Isolation Interference	Displays the percentage of interference by co-isolation within the precursor isolation window. For more information on co-isolation, see "Excluding Peptides with High Levels of Co-Isolation" on page 547.

Table 45. Peptides page parameters (Sheet 5 of 9)

Feature	Description
Ion Inject Time [ms]	Specifies the time used to accumulate ions in the mass spectrometer to reach their target value before they are scanned out, in milliseconds. Low ion injection times are desirable because they indicate high transmission and fast cycle times. Ion injection times that slowly increase might indicate lower transmission, the need for multiplier gain calibration, or both. The Peptides page displays the Ion Inject Time [ms] column by default for every searched spectrum and every PSM for newly processed data.
Peptides Matched	Displays the number of peptide candidates that fall into the precursor mass tolerance window set for the search.
Charge	Displays the charge state of the peptide.
QuanResultID (in MSF files with isotopically labeled quantification results only)	Uniquely identifies a certain quantification result consisting of the quantification values for different quantification channels. For example, if you identify a peptide in both its light and heavy forms, the light and heavy peptides share the same quantification result, so the same identifier appears for both of them in the QuanResultID column.
# Missing Channels (in MSF files with isotopically labeled and area calculation quantification results only)	Displays the number of missing channels in the peptide.
# Single-Peak Channels (in MSF files with for isotopically labeled and area calculation quantification results only)	Displays the number of single-peak channels in the peptide.
Area (in MSF files with for isotopically labeled and area calculation quantification results only)	Displays the area of the peak for the isotopic cluster for the peptide.
Identity High (Mascot only)	Sets a threshold that determines whether peptides are ranked as high confidence when you perform a decoy search to calculate false discovery rates (FDRs). Peptides with an IonScore above this threshold are ranked as high confidence and marked green.

Table 45. Peptides page parameters (Sheet 6 of 9)

Feature	Description
Identity Middle (Mascot only)	Sets a threshold that determines whether peptides are ranked as medium confidence when you do a decoy search to calculate false discovery rates (FDRs). Peptides with an IonScore above this threshold are ranked as medium confidence and marked yellow.
Homology Threshold (Mascot only)	Sets a threshold that determines whether the proposed peptide sequence and the real sequence are considered similar. Peptides with a score higher than this threshold are at least considered similar or homologous.
Quan Channel (in MSF files with isotopically labeled quantification results only)	Displays the channel name of the peptide that is used for quantification.

Table 45. Peptides page parameters (Sheet 7 of 9)

Feature	Description
Quan Info (in MSF files with quantification results only)	Indicates why a peptide was used in quantification or why it was not: • No Quan Values: No isotopic or isobaric labels were detected in the sample. • No Quan Labels: The peptide was not modified by a label. • Redundant: The same MS/MS spectra were used for two separate peptide matches. • No Proteins: The peptide is not associated with a protein, probably because it did not go above the cutoff threshold defined by the search engines, using these parameters: the Protein Cutoff Score parameter in Mascot, the Protein XCorr parameter in SEQUEST, or the Protein Score Cutoff parameter in ZCore. • Unique: The same MS/MS spectra were used for one peptide match, and the peptide belongs to one protein. • Not Unique: The peptide was found in another protein. • Indistinguishable Channels: The peptide does not have amino acids that could have the specified labels or it has the specified labels but has defined channels with the same delta masses, rendering the channels indistinguishable. • Excluded by Method: Peptides or quantification results are excluded from protein quantification according to the Single-Peak/Missing Channels Allowed setting in the quantification method. This setting specifies the allowed number of missing quantification channels or single-peak quantification channels that can be included in valid quantification results in the ratio calculation.

Table 45. Peptides page parameters (Sheet 8 of 9)

Feature	Description
Quan Info, continued	• Inconsistently Labeled: Peptides do not always display the expected amount of isotopic labeling. • Filtered Out: Peptides are not used in the quantification because the applied filters removed them. This column appears only when you have loaded an MSF file containing quantification results.
Quan Usage (in MSF files with quantification results only)	Indicates whether the peptide was used to quantify the protein: • Used: The peptide was used in quantification. • Not Used: The peptide was not used in quantification.
Heavy/Light (in MSF files with precursor ion quantification results only)	Displays the isotopic label type. This column appears only when you have loaded an MSF file containing the results of isotopically labeled quantification.
Ratio columns (in MSF files with reporter ion quantification results only)	Display the corrected ratio of the intensity of the fragmented tag in a sample to the intensity of the fragmented tag in the control sample.
Intensity	Displays the intensity of the precursor ion.
m/z [Da]	Displays the mass-to-charge ratio of the precursor ion, in daltons.
MH+ [Da]	Displays the protonated monoisotopic mass of the peptides, in daltons.
ΔM [ppm]	Displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion.
RT [min]	Displays the retention time when the peptide was observed, in minutes.
First Scan	Displays the number of the first scan used to identify the peptide.
Last Scan	Displays the number of the last scan used to identify the peptide. The last scan might not be the same as the first scan if the workflow enabled grouped scans.

Table 45. Peptides page parameters (Sheet 9 of 9)

Feature	Description
MS Order	Indicates the order of the MS spectrum: • MS1 is one MS scan. • MS2 is an MS/MS scan. • MS3 is an MS^3 scan.
Ions Matched	Displays the percentage of ions matched by the search engine. This column is only displayed when peptides are ungrouped. The Proteome Discoverer application multiplies the ratio in this column by 100 and uses it as a value.
Total Ions	Displays the total possible number of ions.
Spectrum File	Displays the name of the file containing the spectra.
Annotation	Provides a location for you to insert and store notes about the search results, your analysis, and quality results.
pRS Score	Displays a peptide score based on the cumulative binomial probability that the observed match is a random event. The value of the pRS score heavily depends on the data scored, but usually scores above 50 indicate a good PSM.
pRS Probability	Estimates the probability (0–100 percent) that the isoform is correct.
pRS Site Probabilities	Estimates the probability (0–100 percent) of each phosphorylation site being truly phosphorylated. pRS probabilities above 75 percent indicate that a site is truly phosphorylated.
Matched Ions	Displays the number of ions matched by the search engine. This column is only displayed when peptides are ungrouped.

Interpreting Your Results with the Peptide Identification Details View

The Peptide Identification Details view shows the analyzed spectra of the selected peptide sequence on the Peptides page. You can use the Peptide Identification Details view to rapidly gauge how good a match is. This view displays a protein's amino acid sequence color-coded to indicate the confidence level. You can consult Table 48 while choosing your peptide identification details.

When you use the CID and HCD activation types, the fragment ion spectra contain peaks of the fragment ions formed by the cleavage of the peptide bond. When you use the ETD and ECD activation types, the fragment ion spectra contain peaks of the fragment ions formed by the cleavage of the amide-alpha-carbon bond. These fragments are used to determine the amino acid sequence. A fragment must have at least one charge for it to be detected.

The fragment ions produced are identified according to where they are fragmented in the peptide. The a, b, and c fragment ions have a charge on the N-terminal side, and the x, y, and z fragment ions have a charge on the C-terminal side. The subscript next to the letter indicates the number of residues in the fragment ion.[1]

For the fragment ions that the Proteome Discoverer application uses, see Table 46.

- To display the Peptide Identification Details view
- To display different fragment ions in the Fragment Matches pane and the Fragment Spectrum pane
- To display neutral loss ions
- To display the precursor ions
- To display fragment ions according to charge state in the ion table
- To specify the units in which to display the fragment ion masses in the ion table

❖ To display the Peptide Identification Details view

1. In an open MSF report, click the **Peptides** tab.
2. Click the peptide row of interest.
3. Choose **Search Report > Show Peptide ID Details**, or click the **Show Protein/Peptide ID Details** icon,  .

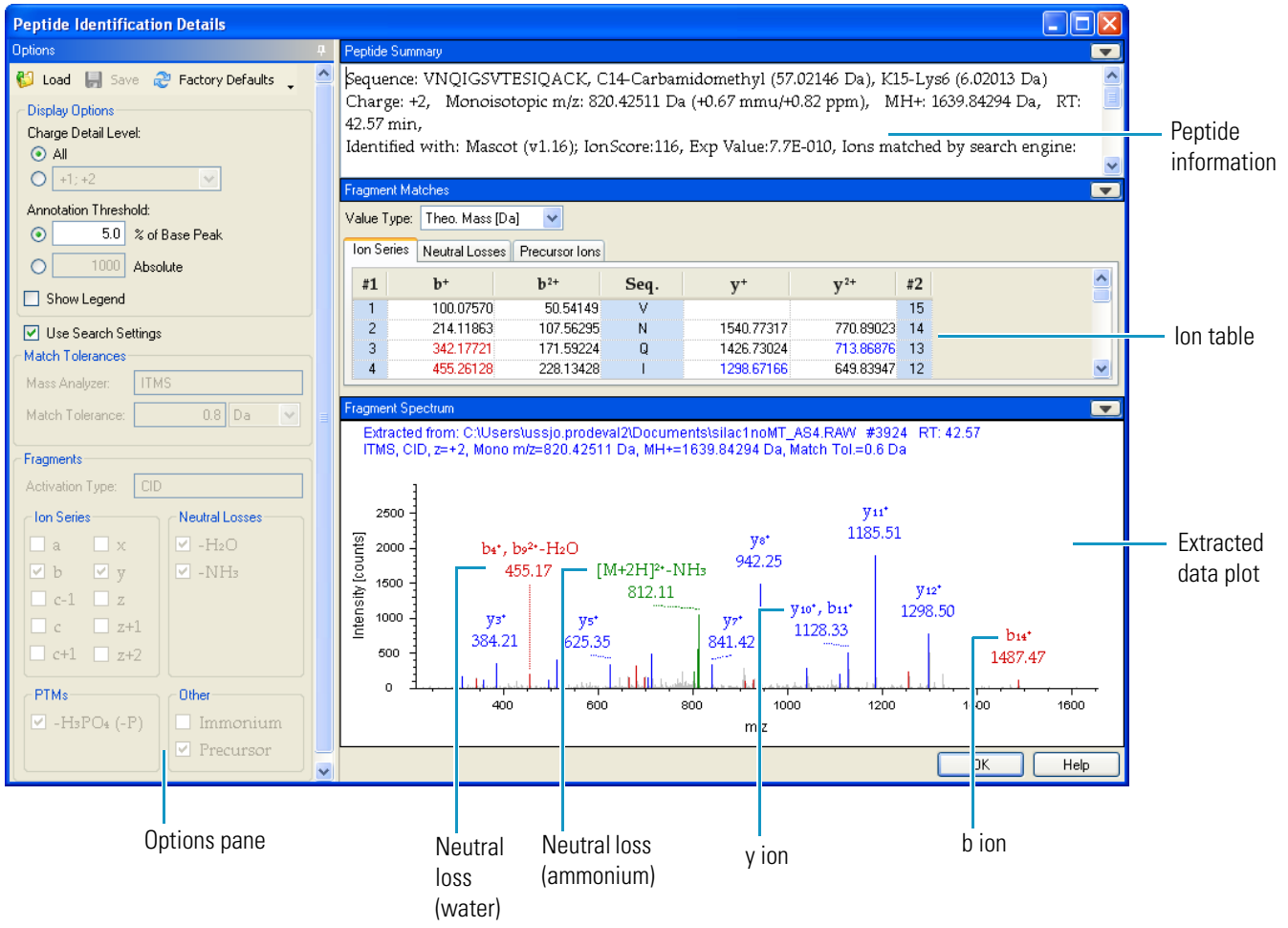
   –or–

   Select a peptide row and double-click a table cell.

   The Peptide Identification Details view appears, as shown in Figure 161.

[1] For more information about fragment ions and nomenclature, see Roepstorff, P. and Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectrum*, **1984**, *11*(11) 601.

Figure 161. Peptide Identification Details view



The Peptide Details Identification view consists of four panes:

- The Options pane on the left is where you choose the data that appears in the Fragment Matches and Fragment Spectrum panes.
- The Peptide Summary pane displays information about the peptide spectra and the conditions of the search.
- The Fragment Matches pane consists of fragment ion tables that display the ion fragments that you selected in the Fragments area of the Options pane.
- The Fragment Spectrum pane displays an extracted plot of the data that you selected in the Options pane.

  The gray lines in the extracted data plot represent unidentified ions in the scan that could be due to chemical “noise” unrelated to the peptide.

For a detailed description of the features of the Peptide Details Identification view, see “Peptide Identification Details View Parameters” on page 310.

You can use the Options dialog box when you choose Tools > Options to change the colors of the ion series and their backgrounds. You can also use this dialog box to change fonts. See "Setting Global Default Fragment Match Options" on page 53.

Note Avoid using the Cambrian Math font in the Peptides Identification Details view, especially in the larger sizes, because it can obscure the data.

Displaying Fragment Ions

The Ion Series page in the Fragment Matches ion table is selected by default. Figure 161 gives an example of this page. The ion series in this table appears in the following default colors:

- Red: a, b, c–1, c, c+1 fragment ions
- Blue: x, y, z, z+1, z+2 fragment ions
- Magenta: immonium ions
- Green: precursor ions
- Yellow background: phosphorylation losses

Note The Proteome Discoverer application automatically chooses the colors of PTMs and neutral loss ions on the basis of the colors of the ion series that they belong to.

❖ To display different fragment ions in the Fragment Matches pane and the Fragment Spectrum pane

1. Select or clear the check boxes for the appropriate ions in the Fragments area of the Options pane.
2. Click the appropriate tab in the ion table.

Table 46 summarizes the fragment ions that appear in the pages of the ion table and in the extracted data plot of the Peptide Identification Details view.

Table 46. Fragment ions (Sheet 1 of 2)

Fragment	Description
Ion series	
a	A ion with charge on the N-terminal side
b	B ion with charge on the N-terminal side
c	C ion with charge on the N-terminal side
c–1	C ion with charge on the N-terminal side, one fewer proton
c+1	C ion with charge on the N-terminal side, one more proton
x	X ion with charge on the C-terminal side
y	Y ion with charge on the C-terminal side

Table 46. Fragment ions (Sheet 2 of 2)

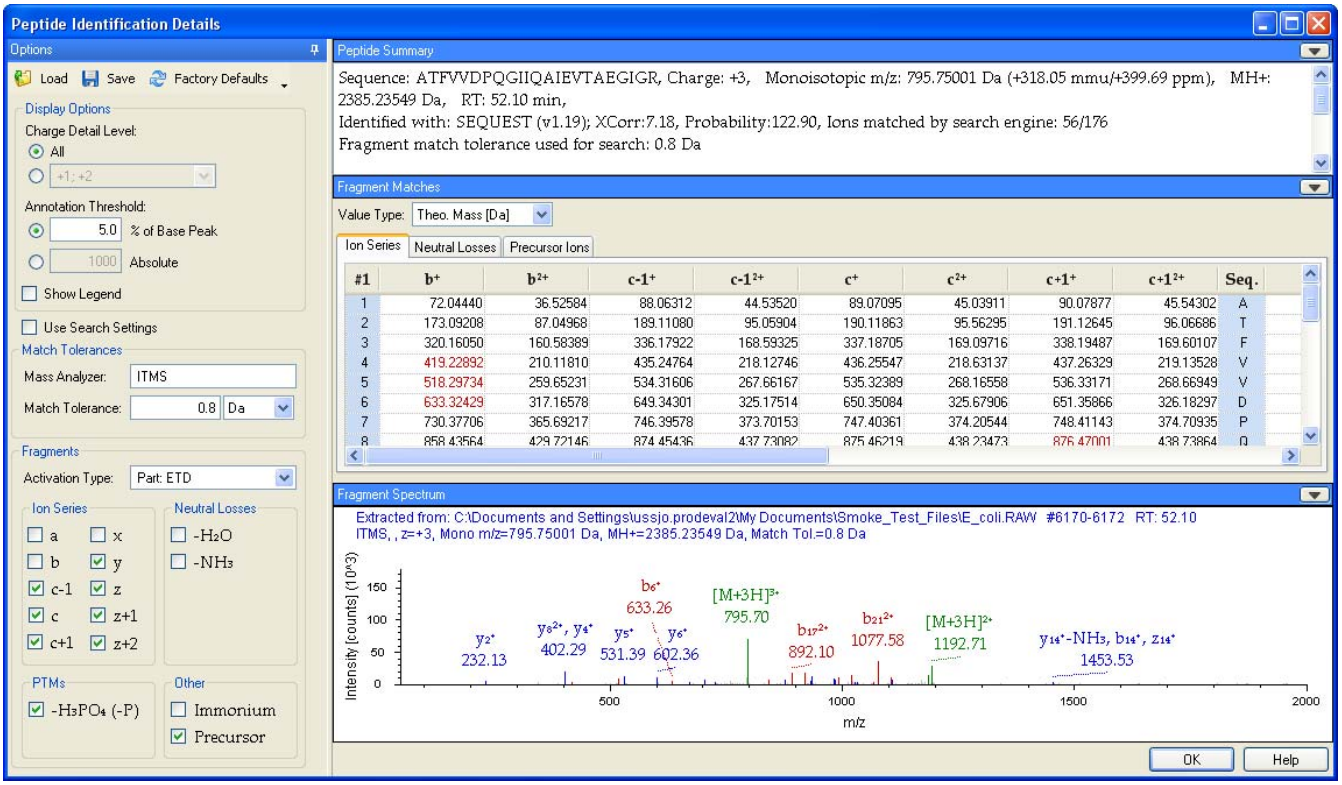
Fragment	Description
z	Z ion with charge on the C-terminal side
z+1	Z ion with charge on the C-terminal side, one more proton
z+2	Z ion with charge on the C-terminal side, two more protons
Neutral losses	
$-H_2O$	Fragment that has lost water (–18 Da)
$-NH_3$	Fragment that has lost ammonia (–17 Da)
PTMs	
H_3PO_4	Phosphorylation loss (designated –P)
Other	
Immonium	Immonium ion
Precursor	Precursor ion

If you select fragment ions in the Fragments area of the Options pane but do not see them displayed in the Fragment Spectrum pane, you might need to take one of the following steps to make them visible:

- If there is not enough room in the Fragment Spectrum pane to display the label text of the fragment ions, you can zoom the spectrum by dragging the left mouse button over the area that you want to enlarge.
- Peaks with a height less than 5 percent of the base peak are not labeled in the Fragment Spectrum pane but are listed in the tables in the Fragment Matches pane. Set the Annotation Threshold option in the Options pane to 0 percent of the Base Peak to see whether the ion labels now become visible.

You might also see fragment ions displayed in the Fragment Spectrum pane, even though you did not select them in the Fragments area of the Options pane. For example, in Figure 162, b ions resulted from the combined methods of ETD and CID. They are displayed in the Ion Series table in the Fragment Matches pane and in the Fragment Spectrum pane, even though they were not selected in the Fragments area of the Options pane.

Figure 162. Unselected b ions in the Fragment Matches and Fragment Spectrum panes

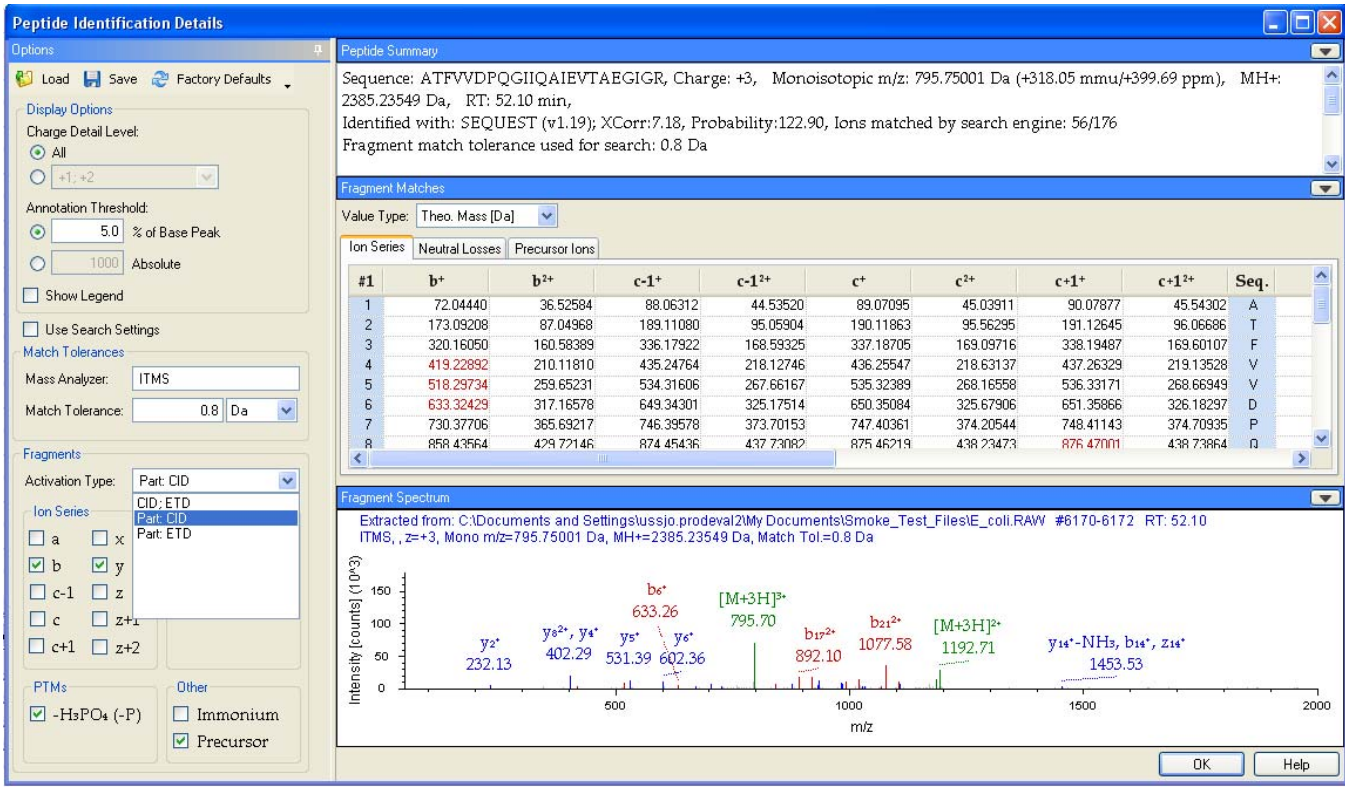


Suppose that you do not want these ions to appear in the data plot in the Fragment Spectrum pane. The Peptide Identification Details view always displays all ions selected in at least one part of the fragmentation method, regardless of which part of the method you selected in the Activation Type box. The current configuration only clears the b ion check box for the ETD part.

❖ To make the b ions invisible

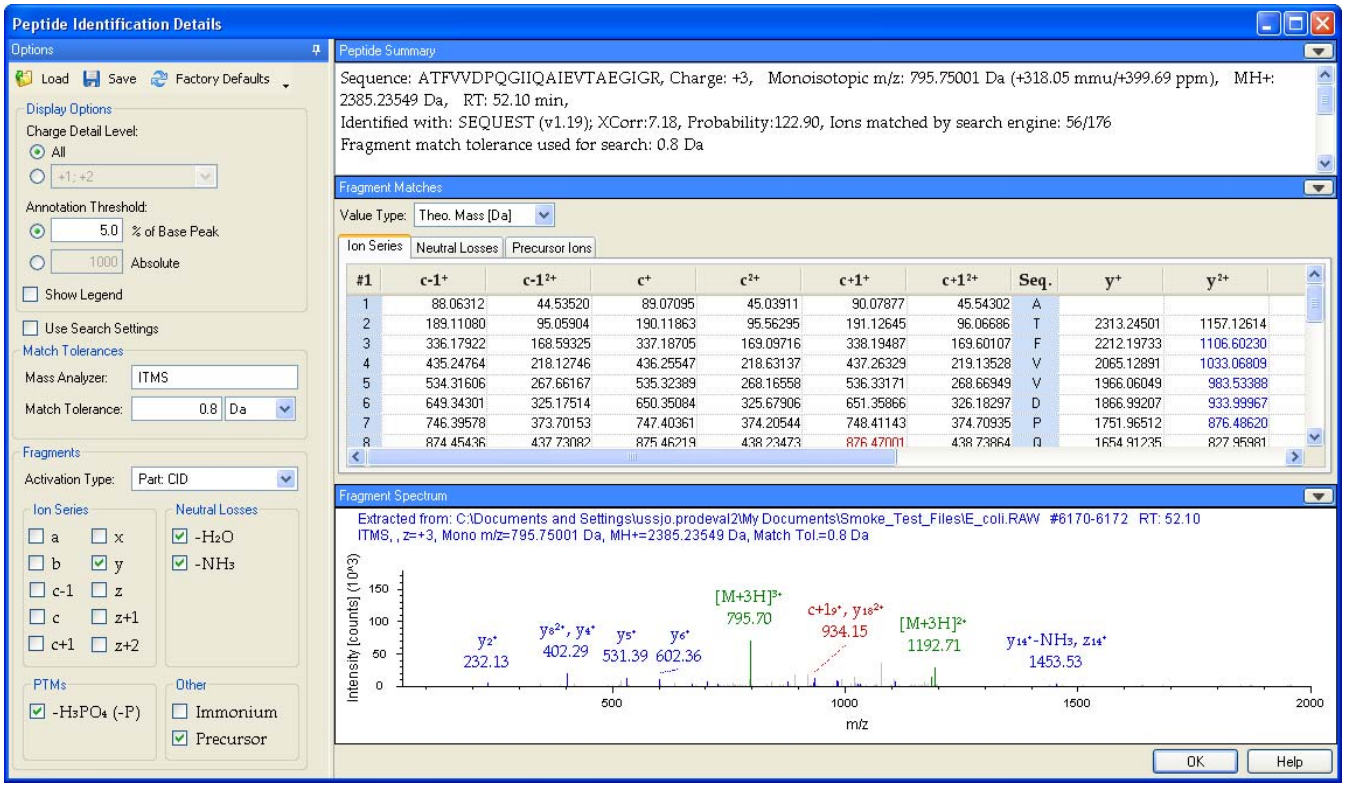
1. Select the CID activation type in the Fragments area of the dialog box, as shown in Figure 163.

Figure 163. Selecting the CID activation type



2. Clear the b ion check box for the CID activation type, as shown in Figure 164.

Figure 164. Clearing the b ions check box



❖ To display neutral loss ions

1. Select the appropriate neutral loss ion check boxes in the Neutral Losses area of the Fragments area.

   Table 47 lists the neutral loss ions that you can select. The Proteome Discoverer application only calculates water neutral loss fragments if side chains contain the potential neutral loss groups of aspartic acid, glutamic acid, serine, or threonine. It only calculates ammonia neutral loss fragments if side chains contain the potential neutral loss groups of lysine, asparagine, glutamine, or arginine.

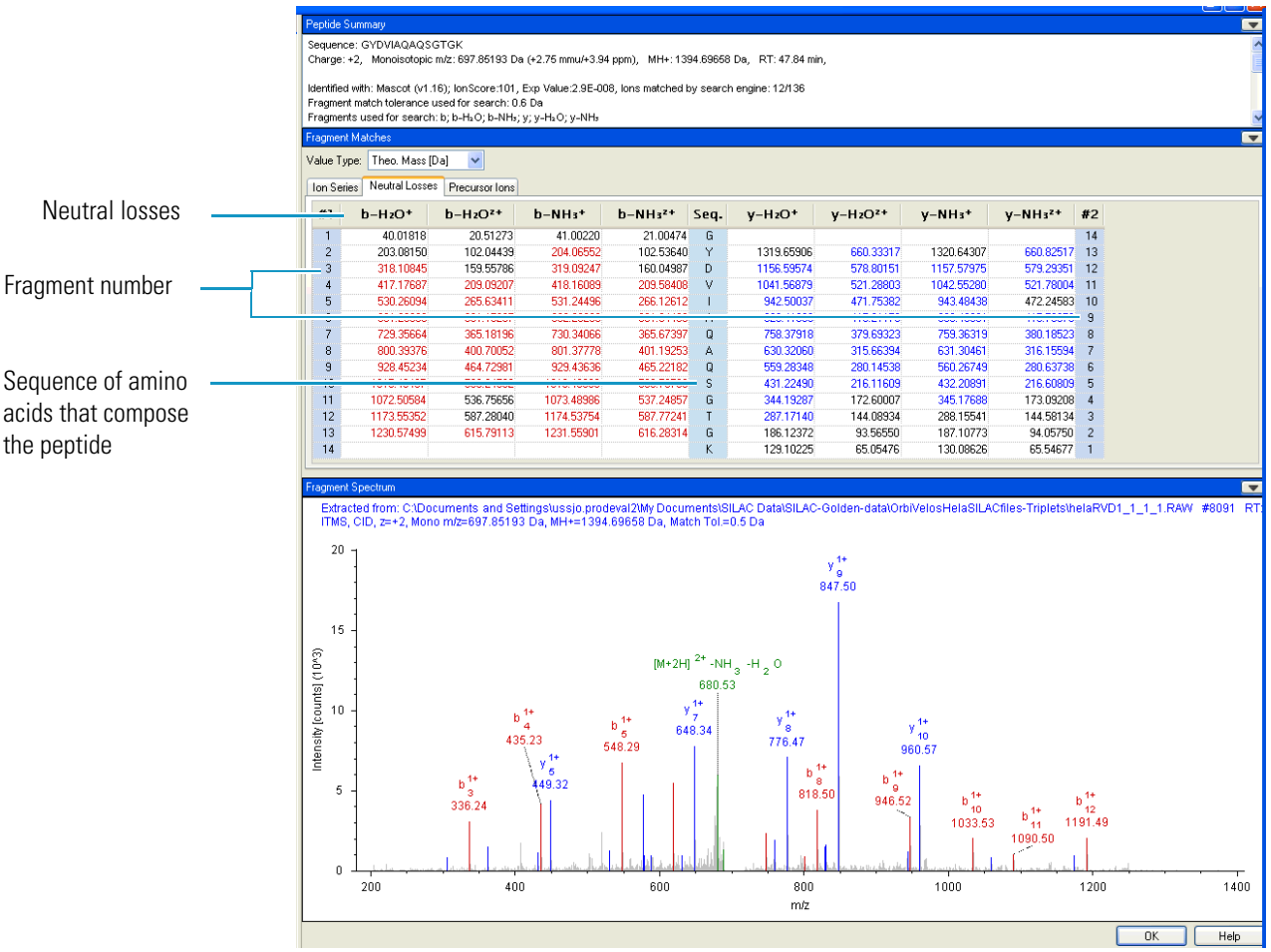
   Table 47. Neutral loss ions

Ions	Description
$-H_20$	Neutral loss ion that has lost water (–18 Da)
$-NH_3$	Neutral loss ion that has lost ammonia (–17 Da)

2. Click the **Neutral Losses** or **Multiple Neutral Losses** tab in the ion table.

Figure 165 shows the Neutral Losses page of the Peptide Identification Details view.

Figure 165. Neutral Losses page of the Peptide Identification Details view

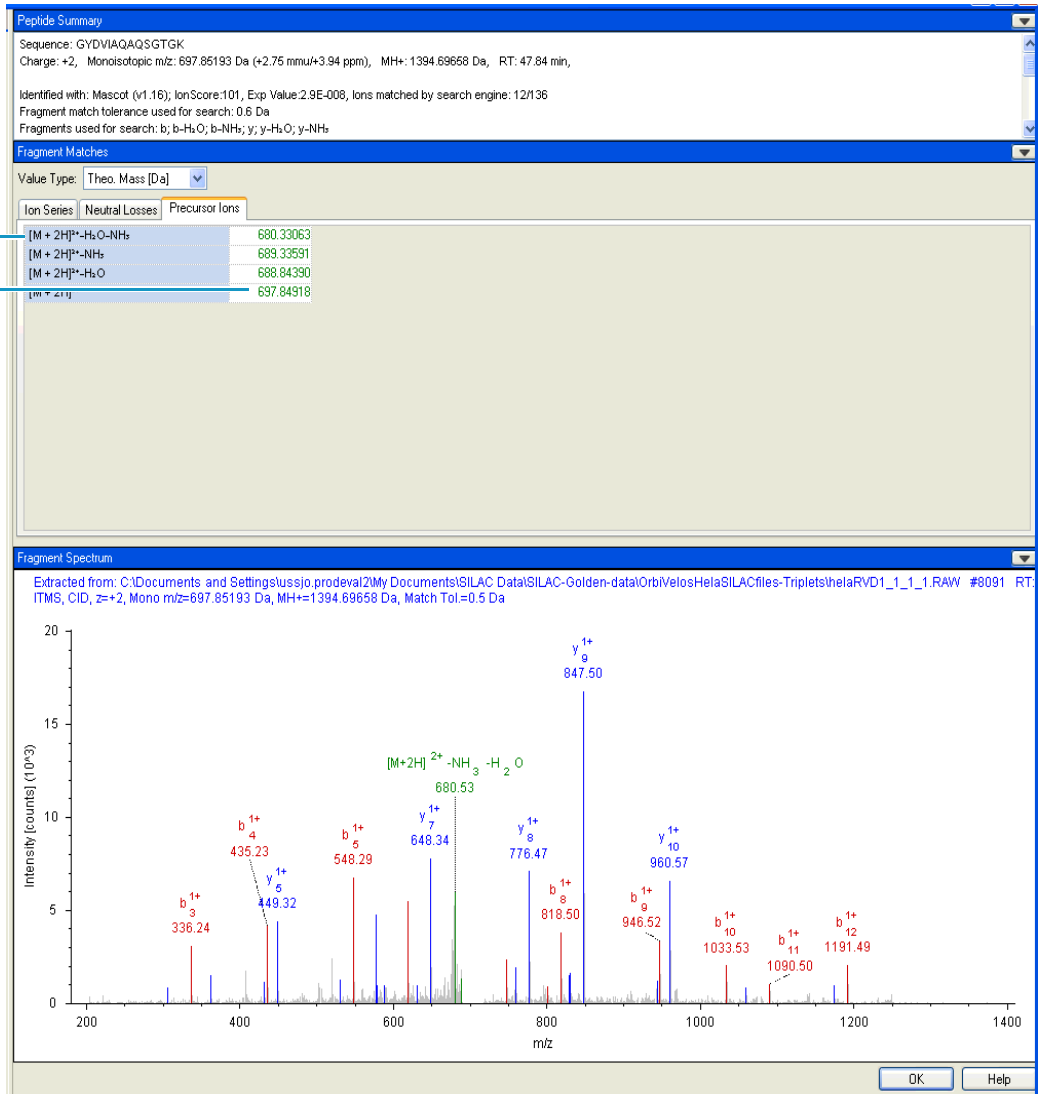


❖ To display the precursor ions

1. Select the **Precursor** check box in the Show Ion Series row.
2. Click the **Precursor Ions** tab in the ion table.

On the Precursor Ions page, the Proteome Discoverer application displays the precursor ions in the peptide and their molecular weights (*m/z*), as shown in Figure 166.

Figure 166. Precursor Ions page of the Peptide Identification Details view



❖ To display fragment ions according to charge state in the ion table

- Select the appropriate state from the Charge Detail Level list above the ion table. To display fragment ions according to charge state in the extracted data plot, select the appropriate state from the Charge Detail Level list above the extracted data plot. Both lists offer the charges available:
  - (Default) All: Displays all fragment ions, regardless of charge.
  - +1: Displays singly charged fragment ions.
  - +2: Displays doubly charged fragment ions.
  - +3: Displays triply charged fragment ions.
  - +4 …: Displays additional charged fragment ions if available.

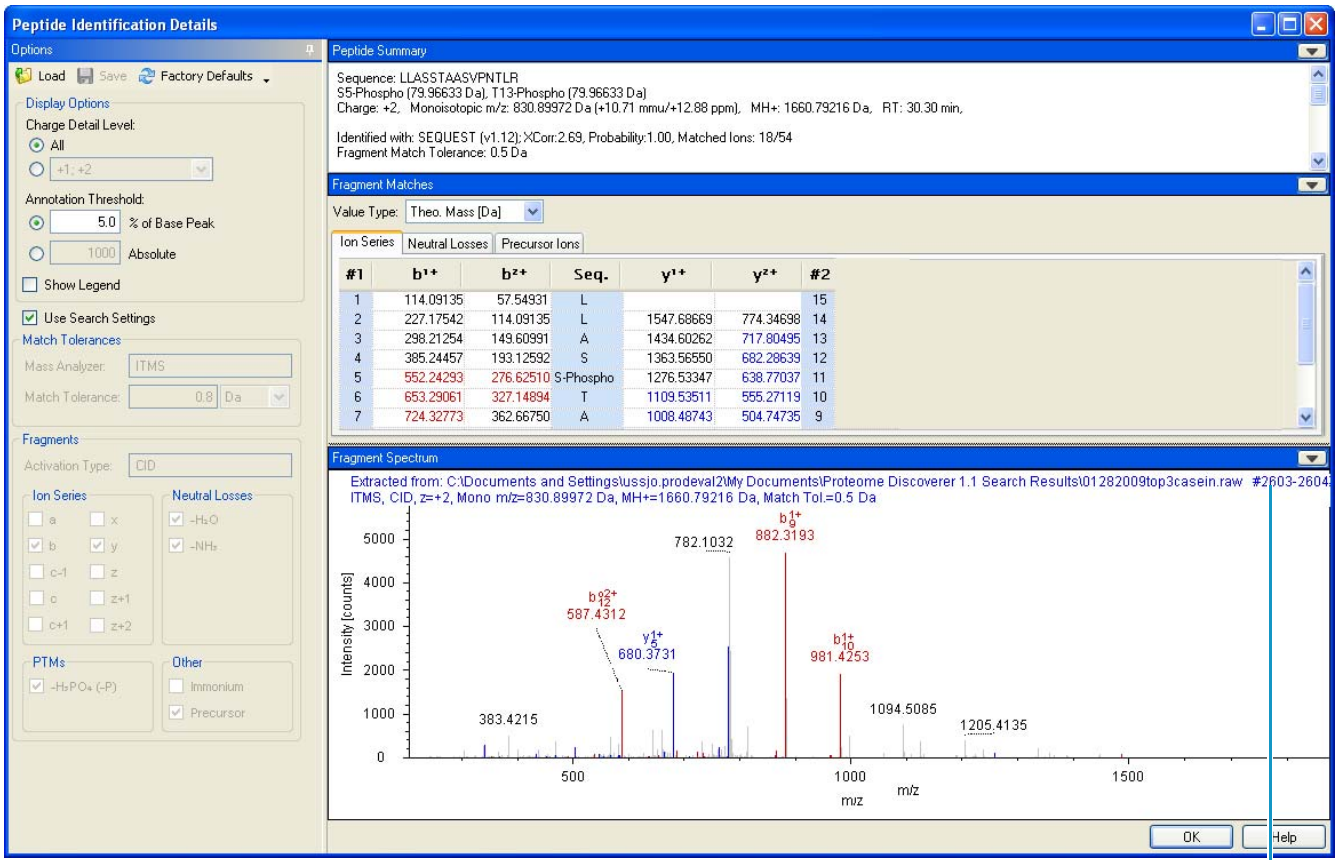
❖ **To specify the units in which to display the fragment ion masses in the ion table**

- Select the appropriate unit from the Value Type list:
  - (Default) Theo. Mass [Da]: Displays the fragment ion masses in daltons.
  - Delta Mass [mmu]: Displays the delta ion masses, in milli-mass units.
  - Delta Mass [ppm]: Displays the delta ion masses, in parts per million.
  - Intensity [counts]: Displays the intensity of the actual ion associated with the peptide fragment, in counts.

Grouping Scans

You can use the Peptide Identification Details view with grouped scans to identify a peptide that has neutral loss and PTMs, such as phosphorylation. These effects might not appear in one type of scan but do appear in another type of scan. For example, Figure 167 shows grouped data (in this case, scans #2603 through #2604) from MS/MS and MS^3 scans from a data-dependent neutral loss experiment. The data from the MS^3 scan shows the neutral loss and PTMs in the peptide, but the data from the MS/MS scan does not.

Figure 167. Grouped data in the Peptide Identification Details view



Peptide Identification Details View Parameters

Table 48 lists the parameters on the pages of the Peptide Identification Details view.

Table 48. Peptide Identification Details view parameters (Sheet 1 of 6)

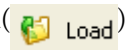
Command or Option	Description
Options	
Load()	Loads the last saved default fragment match settings for the selected mass analyzer/activation type combination.
Save()	Saves the current fragment match settings as the default fragment match settings for the selected mass analyzer/activation type combination.
Factory Defaults()	Loads the original factory fragment match settings for all mass analyzer/activation type combinations.
Charge Detail Level	Selects the appropriate charge state to display. The options are: • (Default) All: Displays all fragment ions, regardless of charge. • +1: Displays singly charged fragment ions. • +2: Displays doubly charged fragment ions. • +3: Displays triply charged fragment ions. • +4 …: Displays additional charged fragment ions if available.
Annotation Threshold	Specifies a peak intensity limit. Peaks below this limit are not annotated. This limit prevents the addition of too many labels to a spectrum, especially if the low-intensity peaks are possibly noise peaks. • (Default) % of Base Peak: Specifies an intensity threshold as a percentage of the base peak in the spectrum. The base peak is the largest peak in a spectrum. • Absolute: Specifies an absolute intensity threshold.
Show Legend	Determines whether a legend consisting of colored lines denoting the displayed ion series appears at the bottom of each spectrum image.
Use Search Settings	Annotates the spectra using the Match Tolerances and Fragments settings used for the search.

Table 48. Peptide Identification Details view parameters (Sheet 2 of 6)

Command or Option	Description
Mass Analyzer	Specifies which mass analyzer was used to obtain the spectrum.
Mass Tolerance	Specifies the mass tolerance. The Proteome Discoverer application annotates the spectrum with ions that it assigns within this mass tolerance. You can use daltons or milli-mass units.
Activation Type	Specifies what type of activation produced the fragments.
Ion Series	Selects which series of fragment ions are displayed in the ion table and in the extracted data plot. By default, the spectrum displays the most relevant ion types for each activation type. See Table 46 for a list of the ions series that you can display. By default, the ion series are displayed in the following colors: • Red: a, b, c–1, c, c+1 fragment ions • Blue: x, y, z, z+1, z+2 fragment ions • Magenta: immonium ions • Green: precursor ions • Yellow background: phosphorylation losses
Neutral Losses	Determines whether water or ammonia neutral losses are displayed in the ion table and in the extracted data plot. The Proteome Discoverer application adds a column for each type of neutral loss ion that you select to the Neutral Losses page. For example, if you select the b and y ions in the Ion Series area and then choose the $-NH_3$ check box, it adds a b$-NH_3$ and a y$-NH_3$ column to the table on the Neutral Losses page. It also adds these ions to the extracted data plot. The Proteome Discoverer application only calculates water neutral loss fragments if side chains contain the potential neutral loss groups of aspartic acid, glutamic acid, serine, or threonine. It only calculates ammonia neutral loss fragments if side chains contain the potential neutral loss groups of lysine, asparagine, glutamine, or arginine.

Table 48. Peptide Identification Details view parameters (Sheet 3 of 6)

Command or Option	Description
PTMs	Displays ions that have lost H_3PO_4 in the ion table and in the extracted data plot. This loss is represented as –P.
Other	Annotate the spectra with immonium ions or precursor ions.
Peptide Summary	
Description of spectra in a scroll box	Displays the following information about the spectra: • The amino acid sequence of the peptide • The PTMs, if any • The charge state of the precursor ion • The mass-to-charge ratio (*m/z*) of the precursor ion, in daltons • The MH+ (the protonated monoisotopic mass), in daltons • The retention time of the precursor ion, in minutes Below this information appears the following: • The search node or wizard used • The XCorr (SEQUEST), IonScore (Mascot), or Probability (ZCore) value, which scores the number of fragment ions that are common to two different peptides with the same precursor mass and calculates the cross-correlation score for all candidate peptides queried from the database. • The probability of a match (for SEQUEST and ZCore) • The number of matched fragments found out of the total number of different potential fragments in the sample • The fragment match tolerance, in daltons

Table 48. Peptide Identification Details view parameters (Sheet 4 of 6)

Command or Option	Description
Fragment Matches	
Value Type	Specifies which unit is used to display the fragment ion masses in the ion table: • (Default) Theo. Mass [Da]: Displays the fragment ion masses in daltons. • Delta Mass [mmu]: Displays the fragment ion masses in milli-mass units. • Delta Mass [ppm]: Displays the fragment ion masses in parts per million. • Intensity [counts]: Displays the intensity of the precursor ion.

Table 48. Peptide Identification Details view parameters (Sheet 5 of 6)

Command or Option	Description
Ion table	• Ion Series page: Displays the calculated fragment ion masses for the best-fit sequence for each specified activation type. Entries in the table are colored if they match peaks in the spectrum within the match tolerance. A, b, and c fragment ions are displayed from the N terminal to the C terminal. X, y, and z fragment ions are displayed from the C terminal to the N terminal. The peptide sequence is listed in the Seq. column in the middle of the table. On the far left and far right of the table are sequential numbers that the Proteome Discoverer application automatically assigns to the fragment ions in the order that the amino acids appear in the fragment from either the N terminus or the C terminus until each amino acid is numbered in the peptide. The table on the Ion Series page in the Peptide Identification Details view contains the same data as the ion table in the Peptide Consensus view, but it is organized differently. • Neutral Losses page: Displays the calculated fragment ion masses for ions that have lost a water or ammonia neutral fragment. Entries in the table are colored if they match peaks in the spectrum within the match tolerance. • Phosphorylation Loss Series page: Displays the calculated fragment ion masses for ions that have lost a H3PO4 fragment. Entries in the table are colored if they match peaks in the spectrum within the match tolerance. • Multiple Neutral Losses page: Displays the calculated fragment ion masses for ions that have lost multiple neutral fragments (H_2O, NH_3, or H_3PO_4). Entries in the table are colored if they match peaks in the spectrum within the match tolerance. • Precursor Ions page: Displays the calculated fragment ion masses for precursor ions. Entries in the table are colored if they match peaks in the spectrum within the match tolerance.

Table 48. Peptide Identification Details view parameters (Sheet 6 of 6)

Command or Option	Description
Fragment Spectrum	
Extracted data plot	Shows the annotated spectrum with all the fragments that matched an expected fragment from in-silico digestion. • Header: Displays the following information: • Name and location of the raw file • Numbers of any grouped spectra • Retention time of the precursor ion • Type of mass spectrometer used • Fragmentation method used • Charge state of the precursor ion • Mass-to-charge ratio of the precursor ion • MH+ (the protonated monoisotopic mass) • Match tolerance • Intensity [counts] (*y* axis): Displays the intensity of the fragment ions, in counts. • *m/z* (*x* axis): Displays the mass-to-charge ratio of the precursor ion.

Interpreting Your Results with the Report Item Distribution Chart

The Report Item Distribution Chart performs the same function for peptides as it does for proteins. For information on this chart, see "Interpreting Your Results with the Report Item Distribution Chart" on page 257.

Interpreting Your Results with the Peptide Consensus View

The Peptide Consensus view displays all major fragment ion series in the spectrum. The fragment ion series includes a, b, c, x, y, and z ions, as well as water and ammonia losses from b and y fragment ions. It allows you to analyze fragment ion coverage, as well as any PTMs present. Use the individual spectral views to evaluate experimental data obtained with a single activation type.

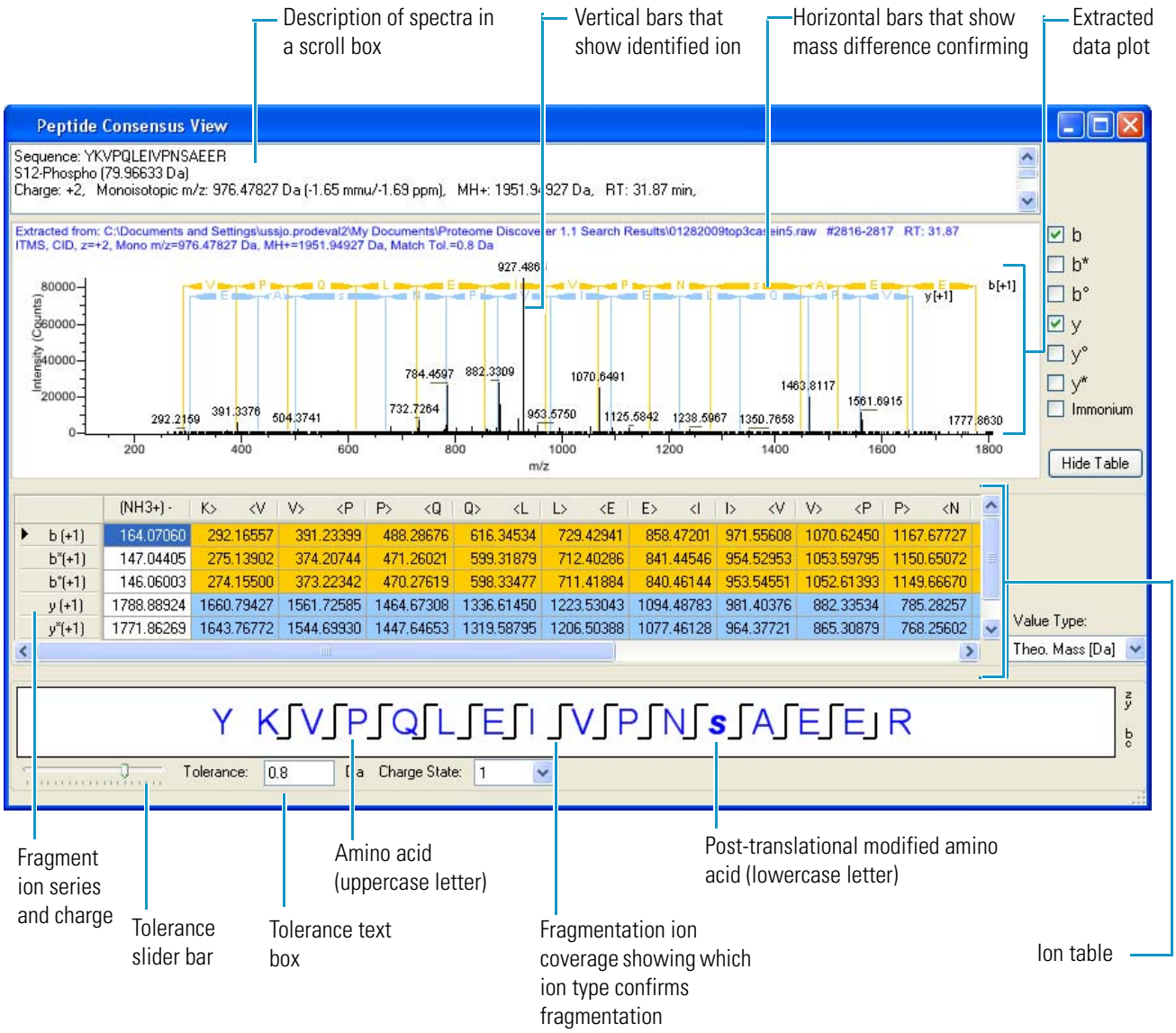
- To display the peptide consensus view
- To display different series of fragment ions in the ion table and the extracted data plot
- To control the visibility of the ion table
- To change the units that will display the fragment ion masses in the ion table
- To display fragment ions according to charge state in the ion table
- To set a tolerance value between the measured and the theoretical fragment masses
- To view information about the post-translational amino acid modifications

❖ To display the peptide consensus view

1. In an open report, click the **Peptides** tab.
2. Click a peptide row.
3. Choose **Search Report > Show Peptide Consensus View**, or click the **Show Peptide Consensus View** icon,  .

   The Peptide Consensus view appears, as shown in Figure 168.

Figure 168. Peptide Consensus view

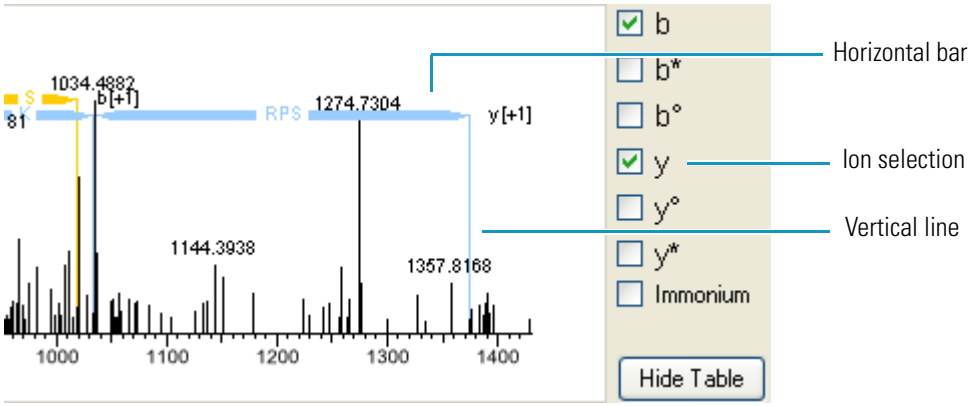


In the Peptide Consensus view, colors have the following significance:

- White indicates that the peptide was not found.
- Gold represents b series fragment ions.
- Blue represents y series fragment ions.
- Orange represents c series fragment ions.
- Green represents z series fragment ions.

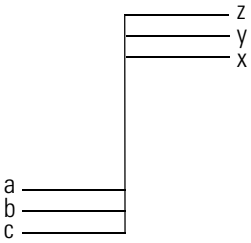
By default, the Peptide Consensus view displays all ion types searched in the analysis. You can remove any of these by clearing the appropriate ion check box on the right, as shown in Figure 169. For a description of these fragment ions, see Table 46.

Figure 169. Selecting the fragment ions to display in the extracted data plot



The ⌋ flag in the fragment ion coverage section of the Peptide Consensus view represents the a, b, and c ions, and the ⌈ flag represents the x, y, and z ions. The a, b, and c ion symbols are distinguishable from one another by their vertical distance from the amino acid, and the same is true for the x, y, and z ions. Figure 170 shows the key to the fragmentation location flags.

Figure 170. Fragmentation location flags



❖ To display different series of fragment ions in the ion table and the extracted data plot

- Select the appropriate fragment ions in the check boxes to the right of the extracted data plot.

❖ To control the visibility of the ion table

- Click **Hide** to make the ion table invisible or **Show** to display the ion table.

❖ To change the units that will display the fragment ion masses in the ion table

- Select the value from the Value Type list:
  - (Default) Theo. Mass [Da]: Displays the fragment ion masses in daltons.
  - Delta Mass [mmu]: Displays the fragment ion masses in milli-mass units.
  - Delta Mass [ppm]: Displays the fragment ion masses in parts per million.
  - Intensity [counts]: Displays the number of counts in the intensity of the fragment ion.

  Or, you can select Delta Mass (in mmu or ppm) and Intensity.

❖ To display fragment ions according to charge state in the ion table

- Select a value in the Charge State list:
  - All: Displays all fragment ions, regardless of charge.
  - (Default) +1: Displays singly charged fragment ions.
  - +2: Displays doubly charged fragment ions.
  - +3: Displays triply charged fragment ions.
  - +4 …: Displays additional charged fragment ions, if available.

❖ To set a tolerance value between the measured and the theoretical fragment masses

- In the Tolerance box, select the existing value, type in the new value, and press ENTER.

–or–

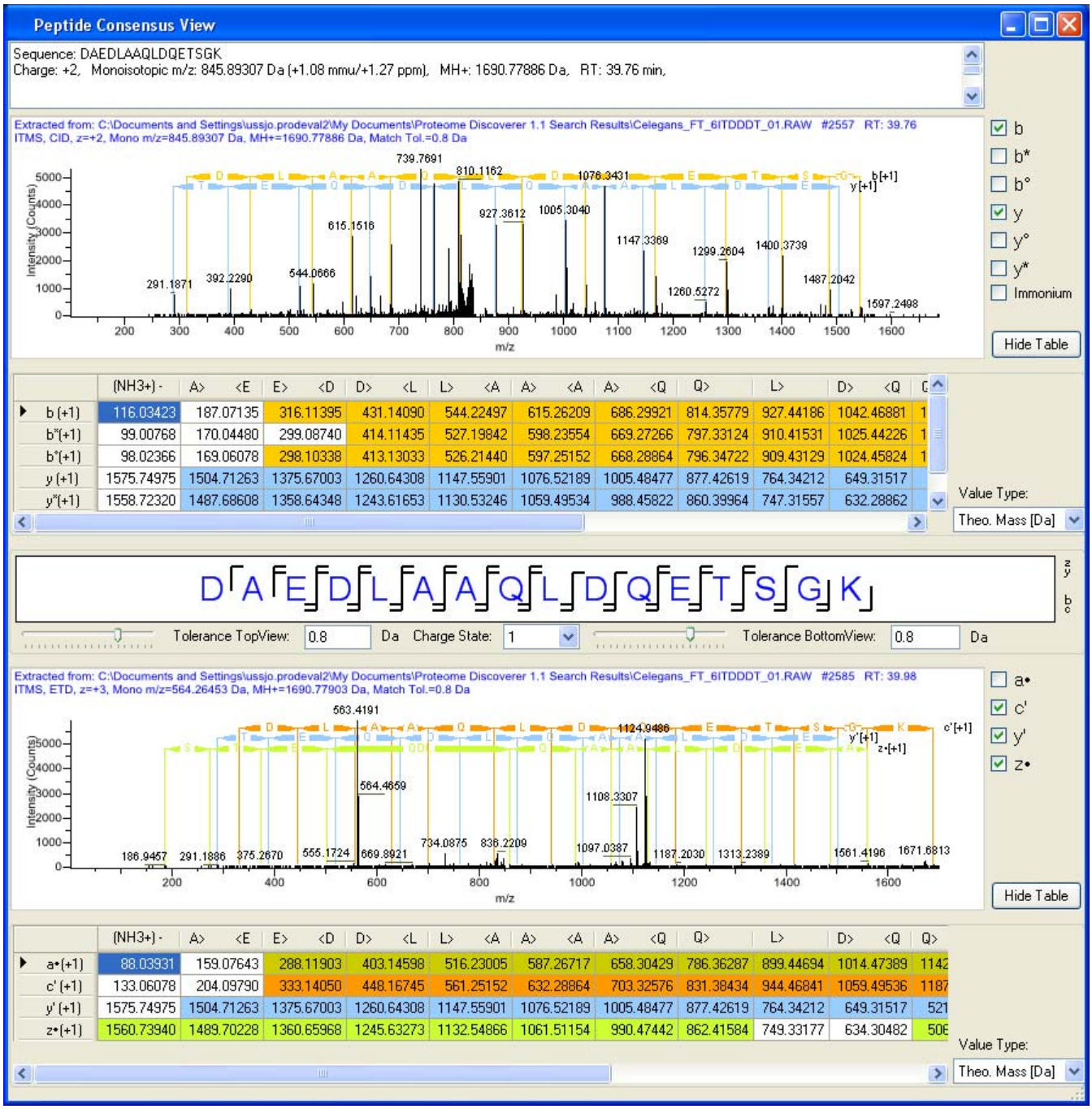
- Use the tolerance slider bar in the lower left corner of the Peptide Consensus view.

❖ To view information about the post-translational amino acid modifications

To obtain information about post-translational amino acid modifications, see "Displaying Post-Translational Modification Information in the UNIMOD Web View" on page 327.

You can use the Dual-Peptide Consensus view to compare two scans to confirm the identity of a peptide. The example in Figure 171 shows the fragment ion masses resulting from two scans from different activation types that both identify the same peptide. This data was captured by alternating CID and ETD scans and then was sorted by time.

Figure 171. Peptide Consensus view from CID and ETD scans confirming same peptide



Peptide Consensus View Parameters

Table 49 lists the parameters on the Peptide Consensus view.

Table 49. Peptide Consensus view parameters (Sheet 1 of 4)

Command or Option	Description
Description of spectra in a scroll box	Displays the following: • Amino acid sequence • Charge state of the precursor ion • Mass-to-charge ratio (*m/z*) of the precursor ion, in daltons • MH+ (the protonated monoisotopic mass), in daltons • Retention time of the precursor ion, in minutes
Horizontal bars	Indicate the mass difference that confirms the presence of amino acids.
Vertical lines	Graphically represent the tolerance between the measured and the theoretical fragment masses. Matches between theoretical and experimental fragment ion masses are indicated by vertical lines, color-coded to indicate the ion type.

Table 49. Peptide Consensus view parameters (Sheet 2 of 4)

Command or Option	Description
Extracted data plot	Displays the MS/MS spectra for any activation type and complexity, from those peptides obtained from bottom-up experiments (approximately 7 to 25 amino acids) to those obtained from middle-down (25 to 60 amino acids) experiments. It contains much of the same data as the extracted data plot of the Peptide Identification Details view, except that it also shows the amino acid sequence at the top of the table. • Header: Displays the name and location of the raw file; the numbers of any grouped spectra; the retention time of the precursor ion, in minutes; the type of mass analyzer used to create the data in the raw file; the fragmentation method used; the charge state of the precursor ion; the mass-to-charge (*m/z*) ratio of the precursor ion, in daltons; the MH+ (the protonated monoisotopic mass), in daltons; and the match tolerance. • Intensity [counts] (*y* axis): Displays the intensity of the precursor ion, in counts. • *m/z* (*x* axis): Displays the mass-to-charge ratio (*m/z*) of the precursor ion.
Ion check boxes	Determines which series of fragment ions are displayed with the horizontal bars. By default, the spectrum displays the most relevant ion types for each activation type. To customize the display to show any type of fragment ions, select the ion check box to add or remove the fragment series to or from the spectrum. For a list of the ions series that you can display, see Table 46.
Hide Table/Show Table	Determines whether the ion table is displayed or hidden.

Table 49. Peptide Consensus view parameters (Sheet 3 of 4)

Command or Option	Description
Ion table	Displays the calculated fragment ion masses for the best-fit sequence for each specified activation type. It displays the theoretical masses in a tabular form. The first column displays the ion types observed, and the remaining columns contain the theoretical masses (by default) for each of the corresponding fragmentation locations for the indicated ion type. The highlighted masses in the ion table represent the peaks that match the measured fragment ion masses within a user-specified tolerance. The ion table header shows the amino acid sequence of the peptide from the C terminal to the N terminal.
Value Type	Specifies the units that will display the fragment ion masses in the ion table: • (Default) Theo. Mass [Da]: Displays the fragment ion masses in daltons. • Delta Mass [mmu]: Displays the fragment ion masses in milli-mass units. • Delta Mass [ppm]: Displays the fragment ion masses in parts per million. • Intensity [counts]: Displays the intensity of the precursor ion in counts.
Fragmentation ion coverage notation	Displays the fragment ion coverage of a specified peptide sequence from a length of the shortest sequence found to 60 amino acids. The fragment ion coverage displays the matched fragment mass according to a user-specified tolerance from one spectrum or two spectra. Only the major fragment ion series per activation type are displayed. • Unmodified amino acids are displayed in capital letters. • Post-translationally modified amino acids are displayed in lowercase letters. To obtain additional information about the modifications, double-click the lowercase letter to display an information card.

Table 49. Peptide Consensus view parameters (Sheet 4 of 4)

Command or Option	Description
Tolerance slider bar	Sets a numeric value for the tolerance between the measured and the theoretical fragment masses, in daltons. The default value is set in the search parameters. The slider bar performs the same function as the Tolerance option.
Tolerance text box	Sets a numeric value for the tolerance between the measured and the theoretical fragment masses, in daltons. The default value is set in the search parameters. This option performs the same function as the Tolerance slider bar. You must press ENTER for the change to take effect.
Charge state	Specifies the charge state of the fragment ions to display in the ion table, extracted data plot, and fragmentation ion coverage notation box: • All: Displays all fragment ions, regardless of charge. • (Default) 1: Displays singly charged fragment ions. • 2: Displays doubly charged fragment ions. • 3: Displays triply charged fragment ions. This option performs the same function as the Charge Level Detail option in the Peptide Identification Details view.

Interpreting Your Results with the Dual-Peptide Consensus View

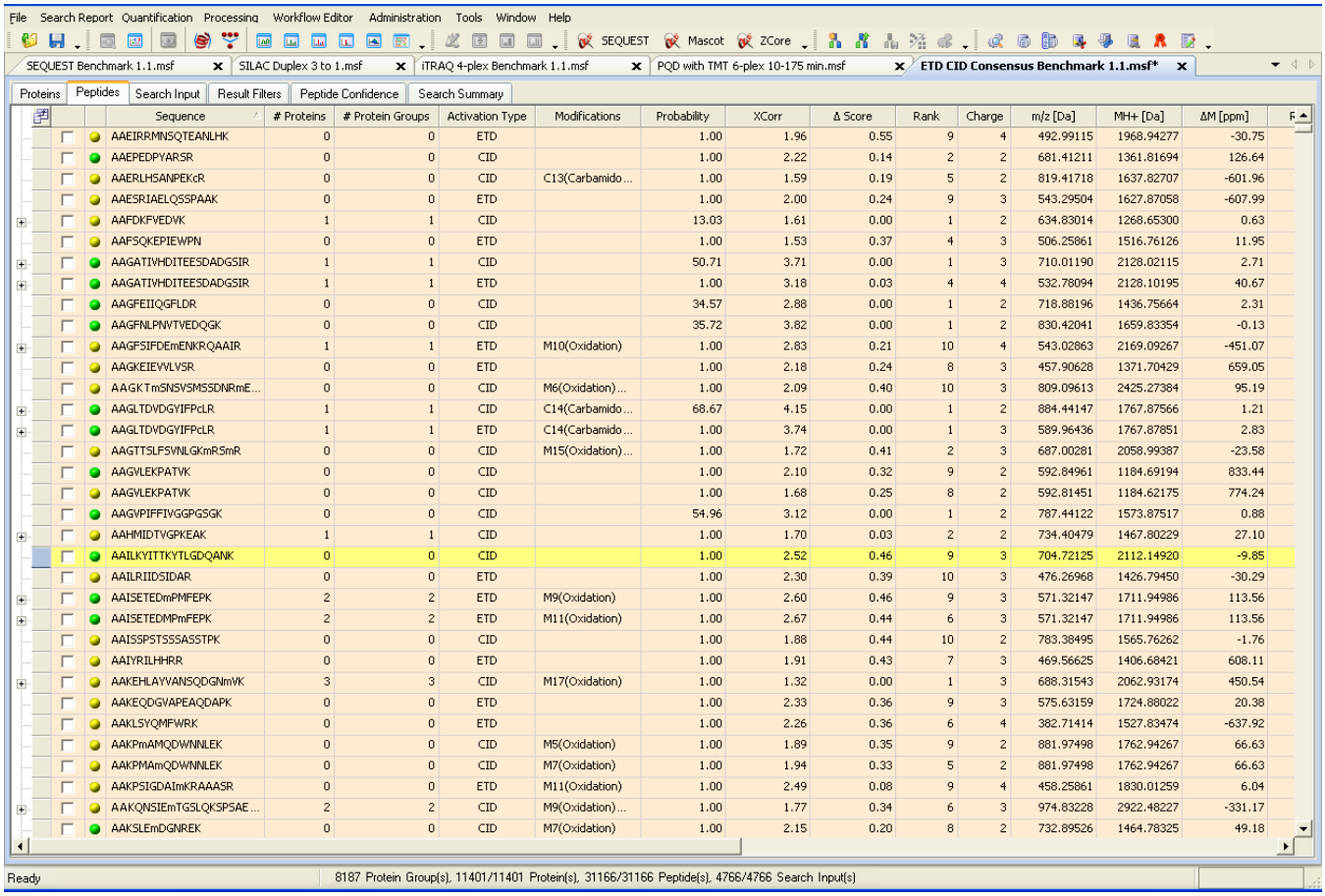
Use the Dual-Peptide Consensus view to compare two files of different activation types for the same best-fit peptide sequence. Use complementary fragmentation methods that produce b- and y-series ions and c- and z-series ions, for example, HCD and ECD, to generate the raw data for the MSF file. In addition, the two confidence markers on the Peptides page must appear in different columns, as shown in Figure 173 on page 325.

❖ To display the Dual-Peptide Consensus view

1. In an open report, click the **Peptides** tab.
2. Ungroup the peptides by right-clicking and choosing **Show Peptide Groups**.

   The peptide view of your search report appears, as shown in Figure 172.

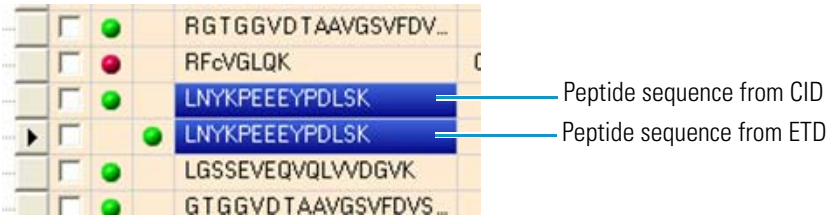
Figure 172. Consecutive CID and ETD scans for the same peptide



3. To sort the data on the Peptides page so that the two identical peptide sequences are contiguous, click the **Sequence** or **RT [min]** column header.

4. Drag the cursor across boxes of identical peptides, as shown in Figure 173.

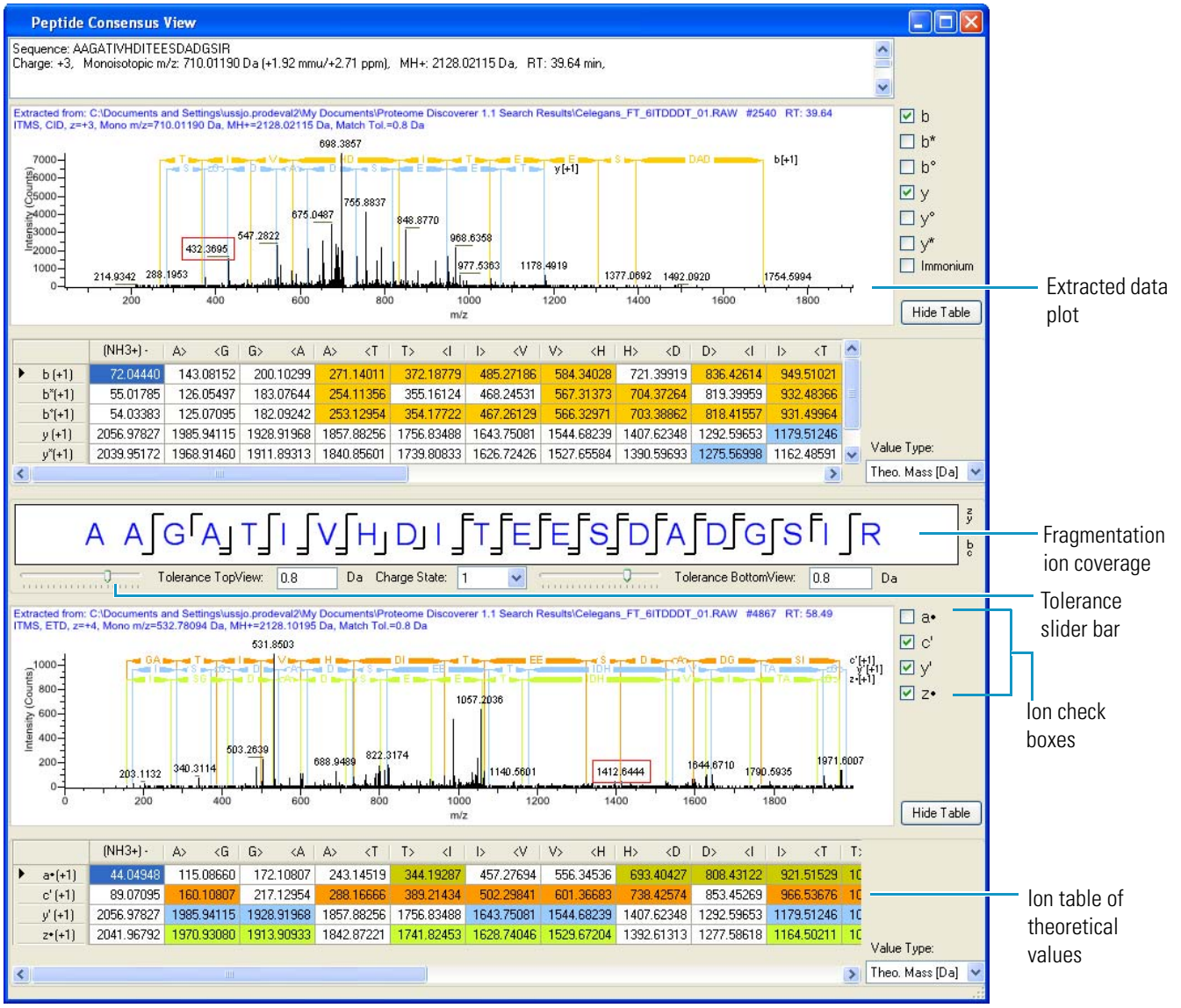
   Figure 173. Identical peptides on the Peptides page



5. Choose **Search Report > Show Peptide Consensus View**, or click the **Show Peptide Consensus View** icon,  .

   The Dual-Peptide Consensus view appears, as shown in Figure 174.

Figure 174. Peptide Consensus View comparing CID and ETD data



In this view, you can see the following:

- Spectra detail as a text header
- Horizontal bars showing mass difference that confirms the presence of amino-acid residues
- Grid cell color indicating positive identification within tolerance and hue-indicated ion series
- Theoretical number values for ions formed by indicated bond breakage in the ion table
- Fragmentation ion coverage showing which ion type confirms fragmentation

6. (Optional) Drag your cursor over spectra region to zoom in. Right-click to reset to initial view.

Displaying Post-Translational Modification Information in the UNIMOD Web View

You can display exact information about an identified PTM to an amino acid by accessing the UNIMOD Web view.

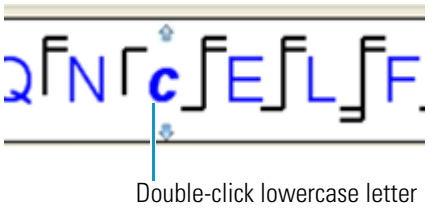
❖ To access the UNIMOD Web view

1. In an open report, click the **Peptides** tab.
2. Click a peptide row.
3. Choose **Search Report > Show Peptide Consensus View**, or click the **Show Peptide Consensus View** icon,  .

   The Peptide Consensus View appears, as shown in Figure 168 on page 317. The multicolored table holds theoretical values.
4. Slide the tolerance slider bar in the lower left corner of the view.

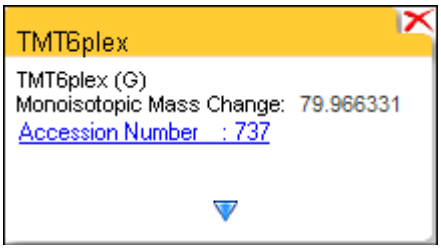
   If the number of highlighted fragment ions does not change, the result is robust.
5. In the fragmentation ion coverage notation, double-click a post-translationally modified amino acid (a lowercase letter), as shown in Figure 175.

   Figure 175. Example of a fragmentation location line indicating the position of an identified fragment

   An Amino Acid Modification Card (AAMC) appears. Each card contains a description of the modification, the mass change data, and a direct link to the corresponding portion of the UNIMOD Web site (http://www.unimod.org).

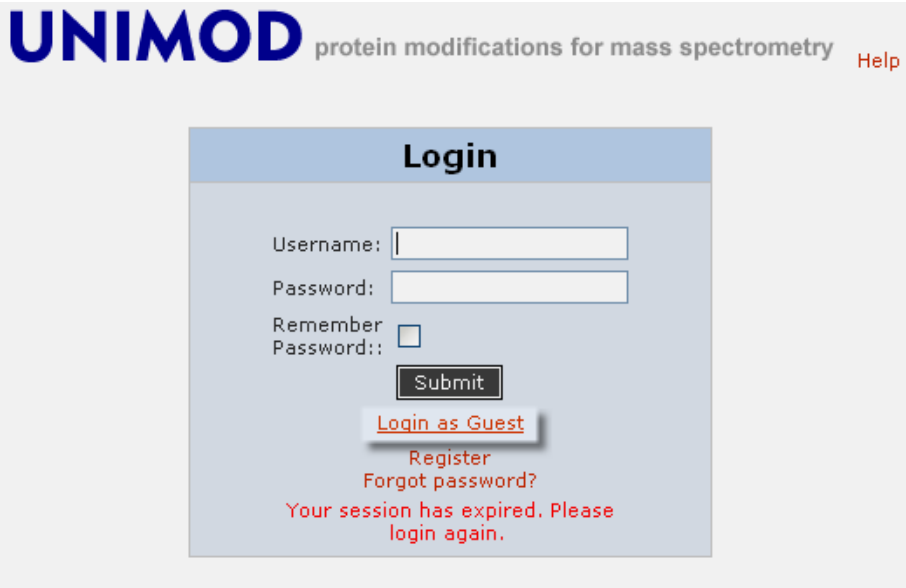
   Figure 176. Amino Acid Modification Card (AAMC)

6. Click the **Accession Number** link.

   A UNIMOD login appears, as shown in Figure 177.

Figure 177. Post-translational modification UNIMOD login



7. Click **Login as Guest**.

The UNIMOD Web view opens, as shown in Figure 178, which displays information about the modification.

Figure 178. UNIMOD Web view

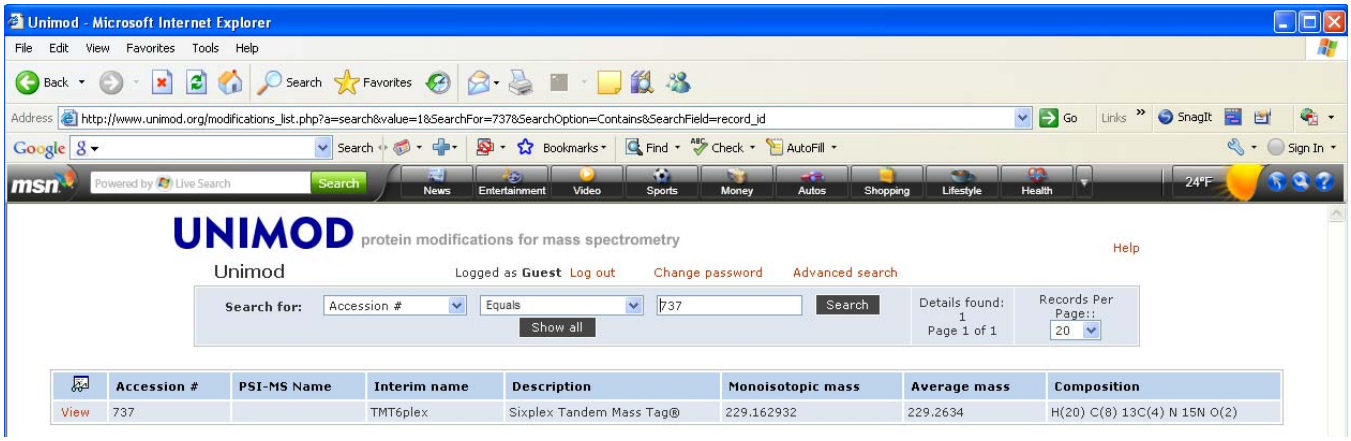


Table 50 describes the parameters in the UNIMOD Web view.

Table 50. UNIMOD Web view parameters

Feature	Description
Amino Acid Modification Card (AAMC)	Provides detailed information about the modified ion. In the fragmentation ion coverage notation, modifications are denoted by a lowercase letter. Click on the lowercase letter to display a popup AAMC. Each card contains a description of the modification, the mass change data, and a direct hot link to the corresponding portion of the UNIMOD Web site (http://www.unimod.org). The AAMC includes the following information: • Amino acid name: Displays the unmodified amino acid name and its one-letter abbreviation. • Modification type: Displays the unmodified amino acid and the modified amino acid names in an abbreviated form. • Mass charge: Displays the delta mass of the current amino acid: • Monoisotopic Mass Change: Displays the delta mass value of the modified and unmodified amino acid. It specifies the monoisotopic mass change value. • Average Mass Change: Displays the delta mass value of the modified and unmodified amino acid. It specifies the average mass change value. • Accession number: Displays a unique identification number for the modification. The accession number is a hyperlink to additional information about the UNIMOD Web site.
Accession #	Displays the accession number automatically assigned to each new entry.
PSI-MS Name	Displays the standardized modification name agreed on by members of the protein modification workgroup of the Proteomics Standards Initiative.
Interim name	Displays the labeling method used by the modification.
Description	Displays a brief description of the modification.
Monoisotopic mass	Displays the monoisotopic mass of the modification.
Average mass	Displays the average mass of the modification.
Composition	Displays the chemical composition of the modification as a delta between the modified and unmodified residue or terminal.

Interpreting Your Results with the Chromatogram View

The Chromatogram view on the Peptides page shows the retention time for all the peptides attributed to a specific peptide group or for a specific peptide.

You can display the base peak chromatogram of the original raw data file in the Chromatogram view. It lists the detected peaks and the qualitative processing results associated with each peak. This chromatogram can provide information about the specific peptides and peak shapes, as well as the intensity of the analytes.

The Chromatogram view is particularly useful in detecting problems in the chromatography or in the ionization.

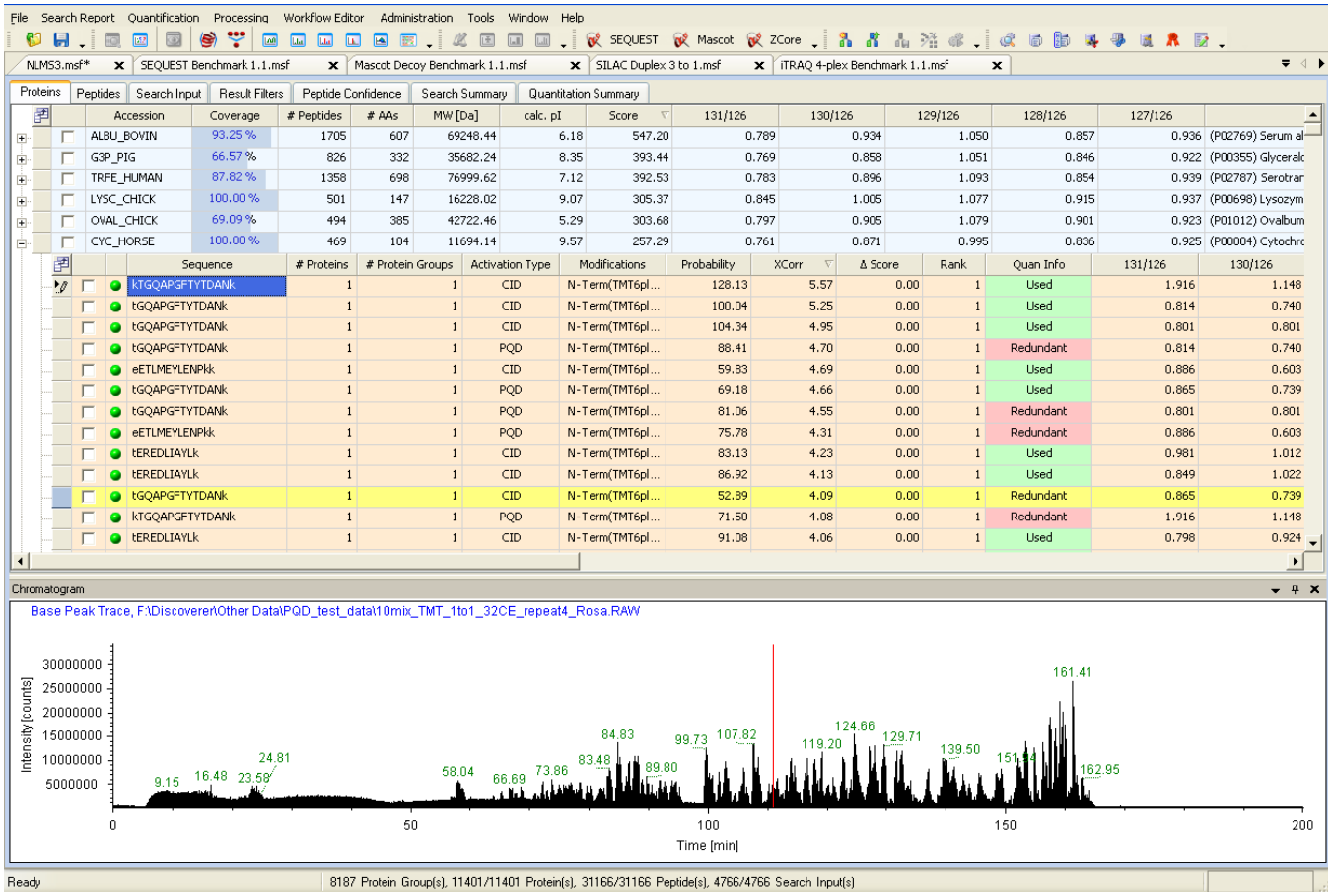
- To display the Chromatogram view
- To display ToolTips
- To zoom in
- To zoom out
- To return to the default view
- To copy the Chromatogram view
- To save the Chromatogram view in an image format

❖ To display the Chromatogram view

1. In an open MSF file, click the **Peptides** tab.
2. Click the row of the peptide of interest.
3. Choose **Search Report > Show Chromatogram View**, or click the **Show Chromatogram View** icon,  .

   The Chromatogram view appears, as shown in Figure 179.

Figure 179. Chromatogram view of selected peptide

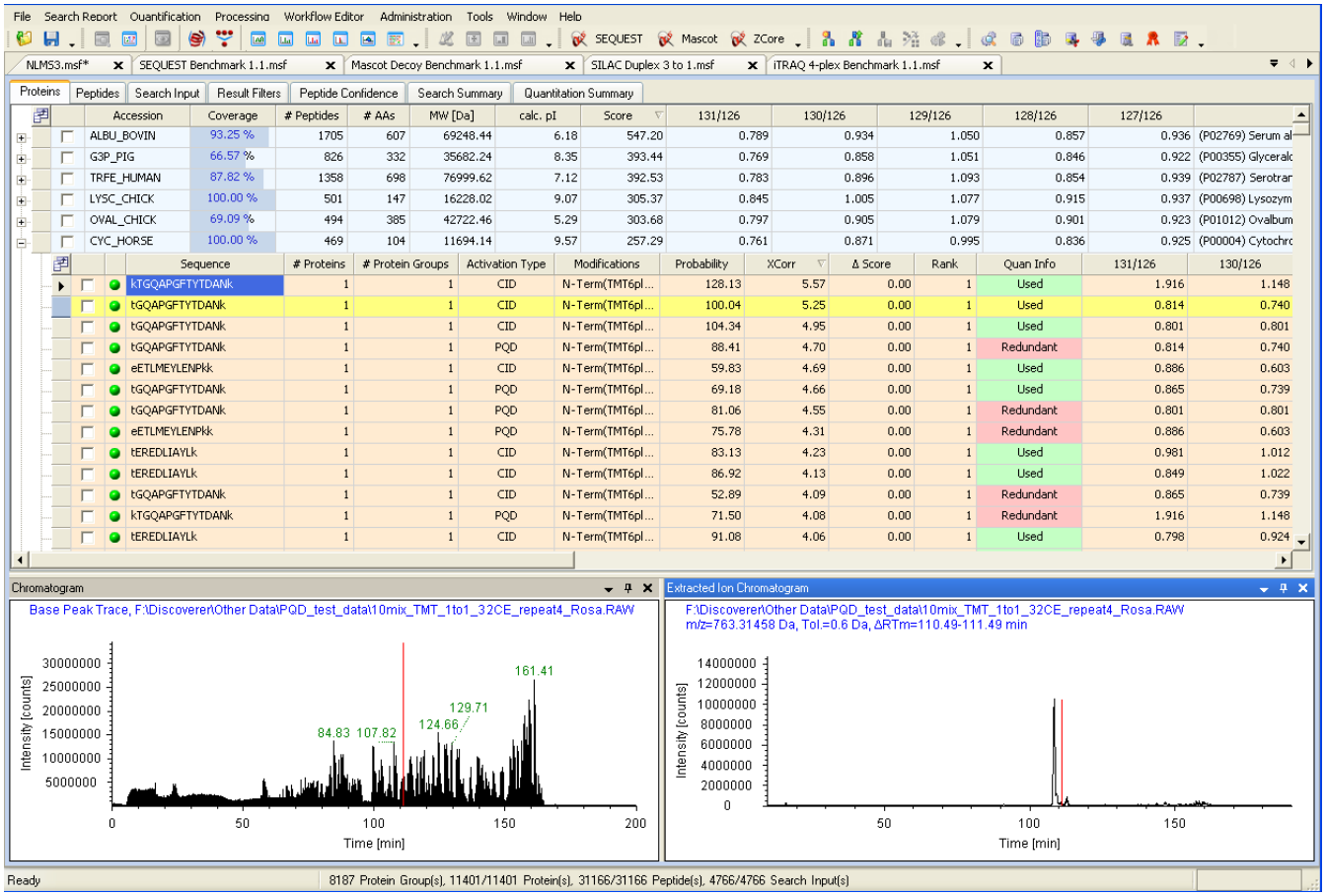


A red line in the chromatogram appears, as shown in Figure 180. Red lines indicate the elution time of the selected peptides.

4. Choose **Search Report > Show Extracted Ion Chromatogram**, or click the **Show Extracted Ion Chromatogram** icon,  .

   The Extracted Ion Chromatogram view appears, as shown in Figure 180. It shows the extracted ion current (XIC) chromatogram of the precursor mass of the selected peptide. For more information about the Extracted Ion Chromatogram view, see "Interpreting Your Results with the Extracted Ion Chromatogram View" on page 340.

Figure 180. Extracted Ion Chromatogram view of selected peptide



The Chromatogram view is interactive and supports multi-level zooming and panning. Use the zoom options to look at the pattern in greater detail.

❖ To display ToolTips

- Right-click the view and choose **Show Position Tooltips**.

  ToolTips now displays the location of the cursor by time and peak height.

❖ To zoom in

- Drag your cursor to the left and select the area to enlarge in size.

❖ To zoom out

- Drag your cursor to the right and select the area to reduce in size.

  –or–

- Right-click the view and choose **Zoom Out** from the shortcut menu.

❖ **To return to the default view**

- Right-click the view and choose **Undo All Zoom/Pan**.

❖ **To copy the Chromatogram view**

- Right-click the view and choose **Copy**.

❖ **To save the Chromatogram view in an image format**

1. Right-click the view and choose **Save As**.
2. In the Save As dialog box, browse to the location where you want to save the file, and type the file name in the File Name box.
3. Click **Save**.

Chromatogram View Parameters

Table 51 lists the parameters on the Chromatogram view.

Table 51. Chromatogram view parameters

Command or Option	Description
Header	Indicates whether the chromatogram is a total ion current (TIC) chromatogram or a base peak chromatogram. The header also displays the name and path of the raw file.
Time [min] (*x* axis for base peak chromatograms)	Displays the elution time for all the peptides in the sample.
Intensity [counts] (*y* axis for base peak chromatograms)	Displays the intensity of the largest peak in the spectrum.
Retention time (*x* axis for total ion current chromatograms)	Displays the retention time of the spectrum.
Intensity [counts] (*y* axis for total ion current chromatograms)	Displays the summed intensity of all peaks in the spectrum.

For additional information about the parameters on the Extracted Ion Chromatogram view, see "Extracted Ion Chromatogram View Parameters" on page 343.

Interpreting Your Results with the Spectrum View

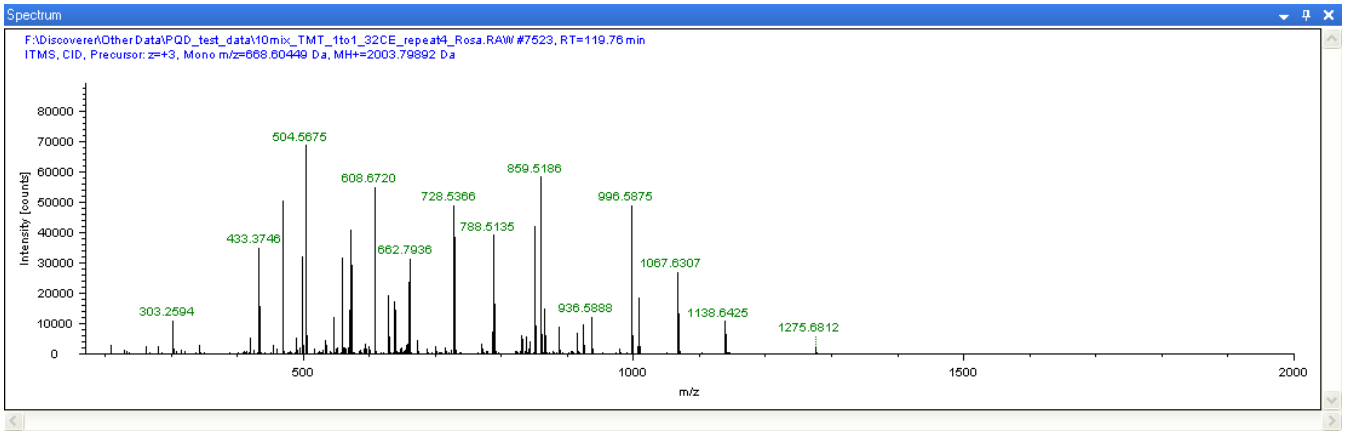
The Spectrum view displays the graphical spectra of the submitted peak list used to search for and identify a peptide. You created the peak list in the Workflow Editor, using the Peak Filters and Spectrum Properties Filter parameters of the Spectrum Selector node. The Spectrum view might differ from the original spectrum in the raw file if you applied spectral preprocessing, such as noise filter or special grouping. Use this view to check the quality of the spectra.

❖ To display a Spectrum view

1. In an open report, click the **Peptides** tab.
2. Click a peptide row.
3. Choose **Search Report > Show Spectrum**, or click the **Show Spectrum** icon,  .

   The Spectrum view appears, as shown in Figure 181. It shows the spectrum used to identify the peptide.

Figure 181. Spectrum view



4. Drag the + cursor to zoom the Spectrum view and examine the ions more closely.

Spectrum View Parameters

Table 52 lists the parameters in the Spectrum view.

Table 52. Spectrum view parameters

Command	Description
Header	Displays the following: • Name and path of the raw file • Numbers of any grouped spectra • Retention time of the precursor ion, in minutes • Type of mass spectrometer used to create the data in the raw file • Fragmentation method used • Charge state of the precursor ion • Mass-to-charge ratio (*m/z*) of the precursor ion, in daltons • MH+ (the protonated monoisotopic mass), in daltons
Intensity [counts] (*y* axis)	Displays the scale of the intensity of the product ions, in counts.
m/z (*x* axis)	Displays the mass-to-charge ratio (*m/z*) of the product ions, in daltons.

Interpreting Your Results with the Fragment Match Spectrum View

The Fragment Match Spectrum view displays the annotated spectrum of the identified peptide. It is the same graph as that shown in the Peptide Identification Details view. However, the Fragment Match Spectrum view allows you to scroll from one peptide to another on the Peptides page and quickly display the annotated spectrum for each peptide.

The matched fragments of the spectrum are colored according to their fragment type, such as blue for b ions and red for y ions.

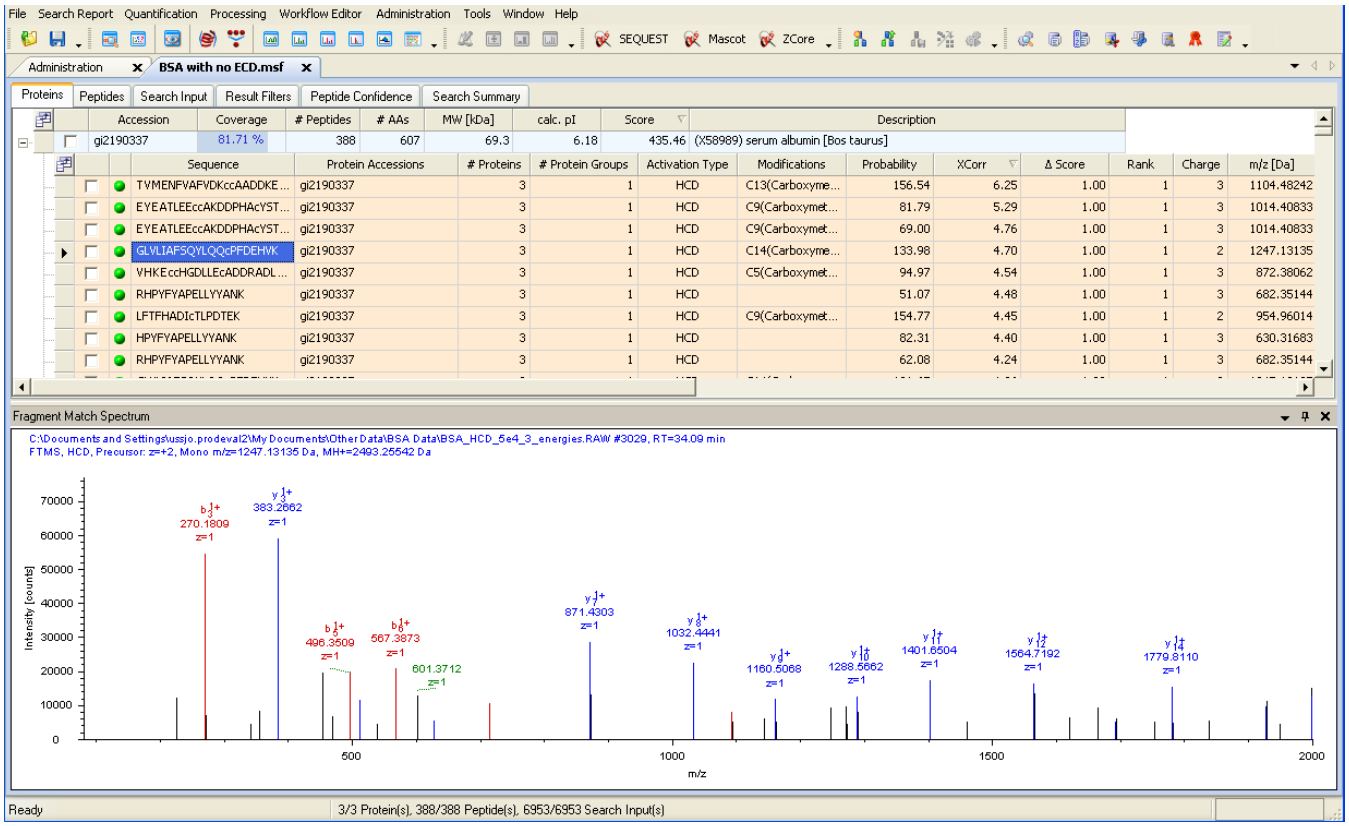
❖ To display the Fragment Match Spectrum view

1. In an open report, click the **Peptides** tab.
2. Ungroup the peptides by right-clicking and choosing **Show Peptide Groups**.
3. Click a peptide row.

4. Choose **Search Report > Show Fragment Match Spectrum**, or click the **Show Fragment Match Spectrum View** icon,  .

   The Fragment Match Spectrum view appears, as shown in Figure 182.

Figure 182. Fragment Match Spectrum view



5. Confirm that matched ions, which are color-coded, are assigned to major fragments.

Note You can use the Fragment Match Options dialog box to change the settings that the application uses to annotate the spectrum with fragment assignments. To display the Fragment Match Options dialog box, right-click on the fragment match spectrum and choose **Fragment Match Options** from the shortcut menu. See "Exporting Search Results in ProtXML Format" on page 596.

Fragment Match Spectrum View Parameters

Table 53 lists the parameters in the Fragment Match Spectrum view.

Table 53. Fragment Match Spectrum view parameters

Parameter	Description
Header	Displays the following: • Name and path of the raw file • Numbers of any grouped spectra • Retention time of the precursor ion, in minutes • Type of mass spectrometer used to create the data in the raw file • Fragmentation method used • Charge state of the precursor ion • Mass-to-charge ratio (*m/z*) of the precursor ion, in daltons • MH+ (the protonated monoisotopic mass), in daltons
Red vertical lines	A-, b-, and c-series ions matched in the spectrum.
Blue vertical lines	X-, y-, and z-series ions matched in the spectrum.
Intensity (counts) (*y* axis)	Displays the scale of the intensity of the product ions.
m/z (*x* axis)	Displays the mass-to-charge ratio of the product ions, in daltons.

Interpreting Your Results with the Precursor Isotope Pattern View

The Precursor Isotope Pattern view displays the isotope pattern of the precursor ion associated with the identified peptide. You can also use this view to assess the abundance and intensity of the precursor ion or to reference the experimentally determined mass of the precursor ion and its isotopes, the isolation window, or other details.

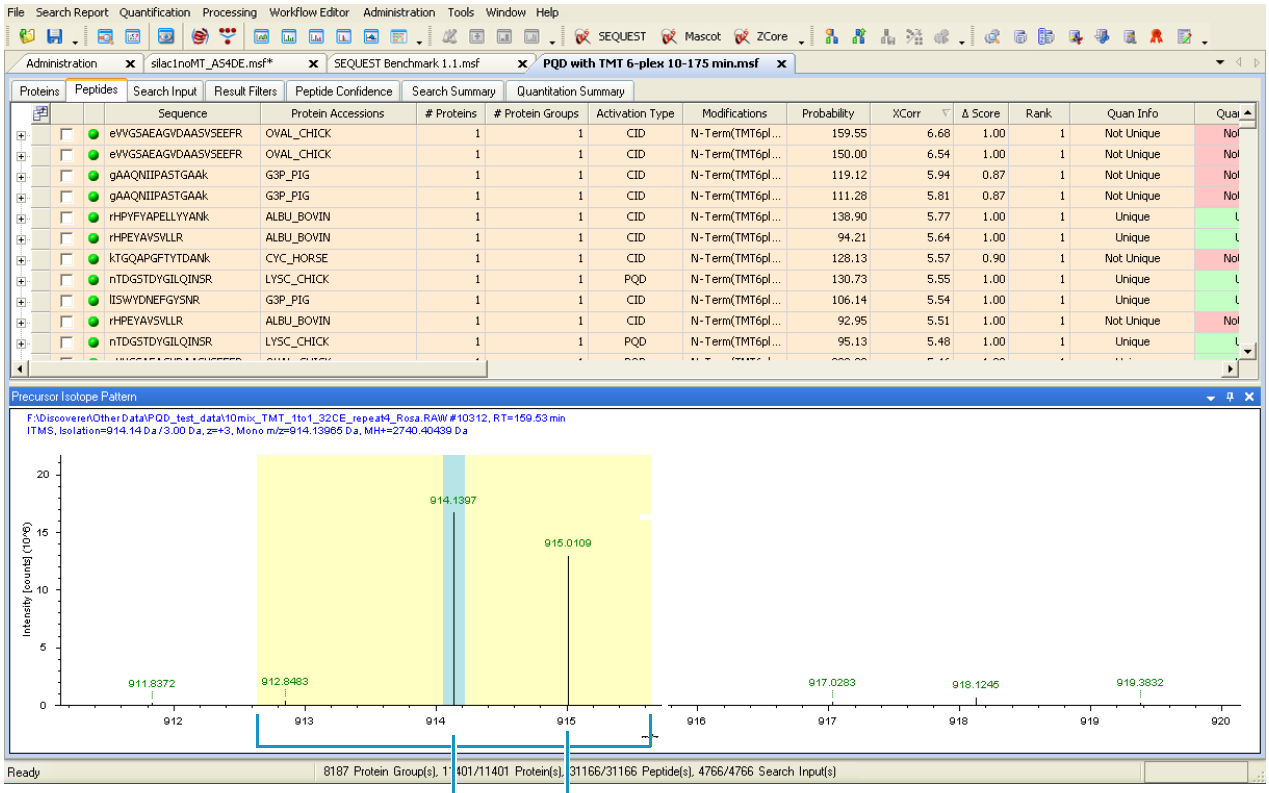
❖ To display the Precursor Isotope Pattern view

1. In an open report, click the **Peptides** tab, or click a peptide beneath a protein on the Proteins page.
2. Ungroup the peptides by right-clicking and choosing **Show Peptide Groups**.
3. Click the appropriate peptide row.

4. Choose **Search Report > Show Precursor Isotope Pattern**, or click the **Show Precursor Isotope Pattern** icon,  .

   The mass spectrum of the precursor ion appears, as shown in Figure 183.

Figure 183. Precursor Isotope Pattern view

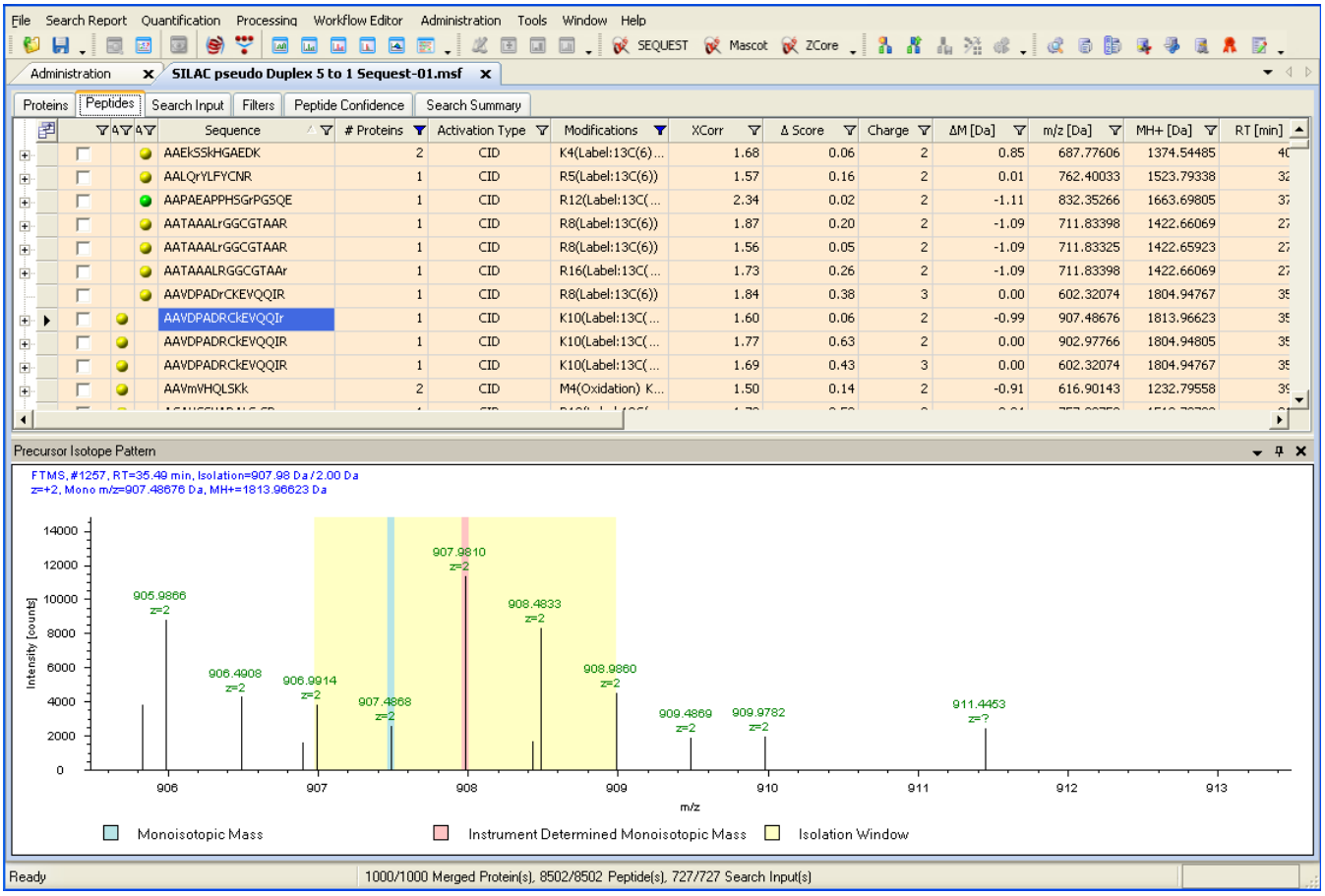


This view has the following components:

- The yellow region displays the isolation width of the instrument, which is the portion of an MS scan that was further subjected to an MS/MS scan.
- A red line indicates the monoisotopic precursor mass-to-charge ratio (*m/z*) determined by the instrument during acquisition.
- A blue line marks the calculated monoisotopic precursor mass and represents the re-evaluated monoisotopic mass-to-charge ratio (*m/z*) value for the detected peptide.

Use the Precursor Isotope Pattern view to verify that the correct monoisotopic mass has been calculated, since the first isotope of a peptide might not be the most intense ion. For example, in Figure 184, the blue line represents the monoisotopic mass that the Proteome Discoverer application calculated, and the red line represents the monoisotopic mass that the mass spectrometer calculated. When the spectrometer selects the correct monoisotopic mass, these two lines should line up. However, in Figure 184, the two lines are divergent, indicating that the mass spectrometer picked the wrong monoisotopic mass.

Figure 184. Precursor Isotope Pattern view showing incorrect monoisotopic mass calculated by the mass spectrometer



Precursor Isotope Pattern View Parameters

Table 54 lists the parameters in the Precursor Isotope Pattern view.

Table 54. Precursor Isotope Pattern view parameters

Parameter	Description
Header	Displays the following: • Name and path of the raw file • Numbers of any grouped spectra • Retention time of the precursor ion, in minutes • Type of mass spectrometer used to create the data in the raw file • Width of the isolation window, in daltons • Charge state of the precursor ion • Mass-to-charge ratio (*m/z*) of the precursor ion, in daltons • MH+ (the protonated monoisotopic mass), in daltons
Intensity [counts] (*y* axis)	Displays the scale of intensity of the precursor scan.
m/z (*x* axis)	Displays the mass-to-charge ratio (*m/z*) of the precursor scan.

Interpreting Your Results with the Extracted Ion Chromatogram View

The Extracted Ion Chromatogram (XIC) view displays the part of the intensity that can be attributed to a certain component—that is, a single mass with a given tolerance. To create an Extracted Ion Chromatogram (XIC) view, the Proteome Discoverer application extracts from every scan that fits your scan filter criteria in the raw file—for example, all MS1 scans—the summed intensity from every mass peak that lies inside the specified mass range, or mass tolerance around a center mass, in each scan. It then plots this extracted intensity in a chromatogram display: along the *x* axis, it displays the retention time of each considered scan, and along the *y* axis, it displays the extracted intensity.

A standard (total ion current (TIC) or base peak) Chromatogram view shows the summed signal from all components with a contribution at a given retention time. The Extracted Ion Chromatogram view shows only the part of the intensity that can be attributed to a certain component—that is, a single mass with a given tolerance. When the application detects a chromatographic peak in this retention-time range, it shades the peak in blue. The red line shows when the MS/MS spectrum was taken.

You can use this view to assess the chromatographic peak shape of the associated precursor and to reference the elution time of the identified peptide.

Use the Extracted Ion Chromatogram graph to determine answers to such questions as the following:

- Does the MS/MS spectrum display poor signal-to-noise ratios due to low abundance of the peptide?
- Was the MS/MS triggered too early (at the start of the peak) or too late (at the end of the peak)?
- Where is the peptide eluting in the chromatogram? Is it as expected?
- Is the shape of the chromatogram as expected?

❖ To display the Extracted Ion Chromatogram view

1. In an open report, click the **Peptides** tab.

   The Peptides page of your search report appears.

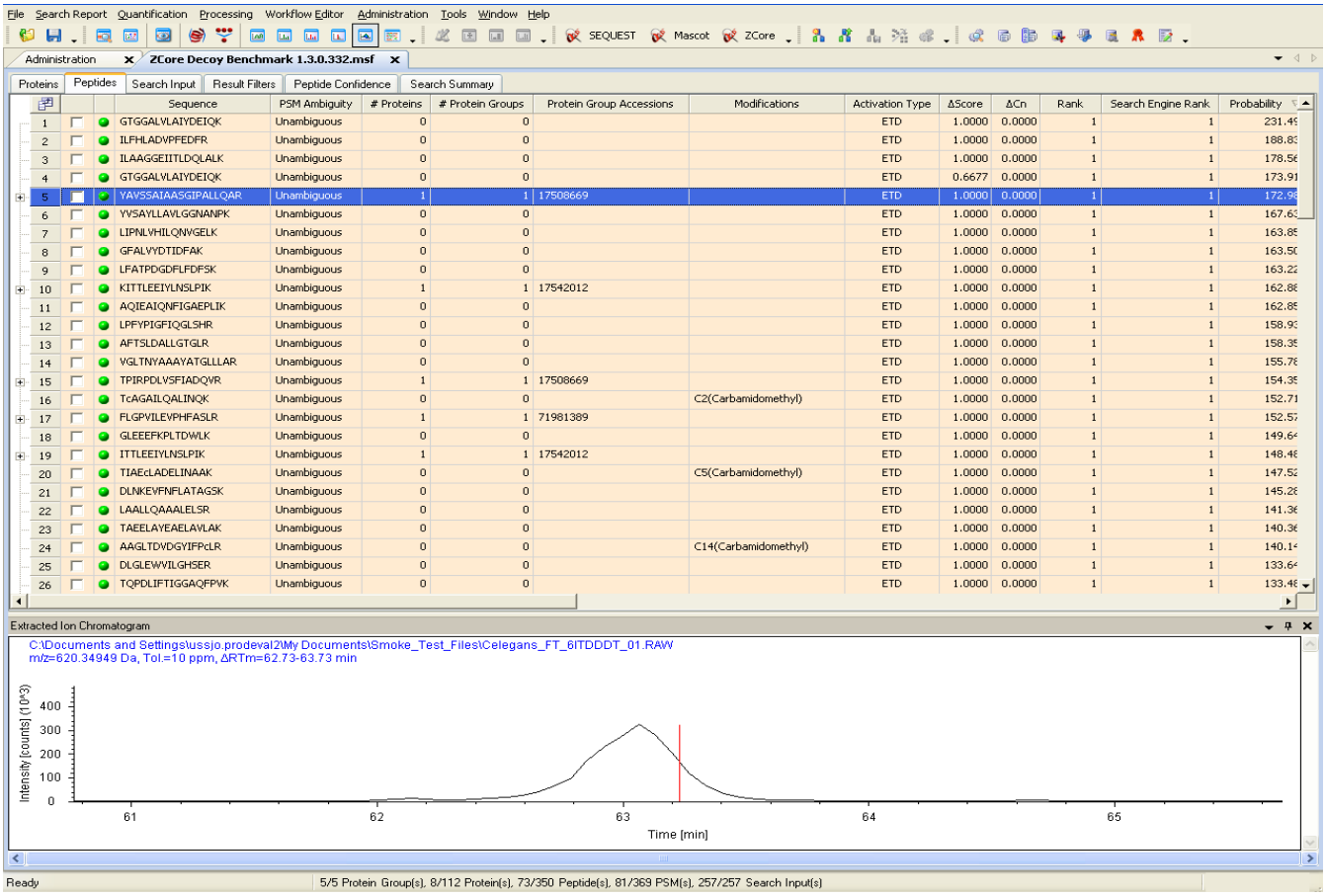
2. Ungroup the peptides by right-clicking and choosing **Show Peptide Groups**.
3. Click the peptide row of interest.
4. Choose **Search Report > Show Extracted Ion Chromatogram**, or click the **Show Extracted Ion Chromatogram** icon,  .

   The Extracted Ion Chromatogram view appears, as shown in Figure 185.

Note The raw file must be in the same folder as the MSF report for the Proteome Discoverer application to construct the Extraction Ion Chromatogram.

Note When you open the Extracted Ion Chromatogram view in a results report containing large amounts of data, it might take a few seconds for the extracted ion chromatogram to display.

Figure 185. Extracted Ion Chromatogram view



5. To magnify a particular peak, drag your cursor over the region of interest.

Extracted Ion Chromatogram View Parameters

Table 55 describes the parameters on the Extracted Ion Chromatogram view.

Table 55. Extracted Ion Chromatogram view parameters

Command or Option	Description
Header	Displays the following information about the raw file and the processing: • The name and path of the raw file. • The mass-to-charge ratio (*m/z*) of the precursor ion, in daltons. • The mass tolerance that was used to extract the extracted ion chromatogram from the raw file, in daltons. • The delta retention time, which gives the retention-time range when the chromatographic peak for a component is likely to occur, in minutes, for example, 60.65–61.64 minutes. When the application detects a chromatographic peak in this retention-time range, the peak is shaded in blue.
Red vertical line	Shows when the MS/MS spectrum was taken.
Intensity [counts] (*y* axis)	Displays the intensity of the precursor ion.
Time [min] (*x* axis)	Displays the elution time of the peptide.

Interpreting Your Results with the Peptide Group Members View

The Peptide Group Members view shows which individual PSMs are included in a peptide group when you enable peptide grouping. The Proteome Discoverer application groups peptides by default. When you step through the Peptides page with the cursor, the Peptide Group Members view stays synchronized with the selected peptide group.

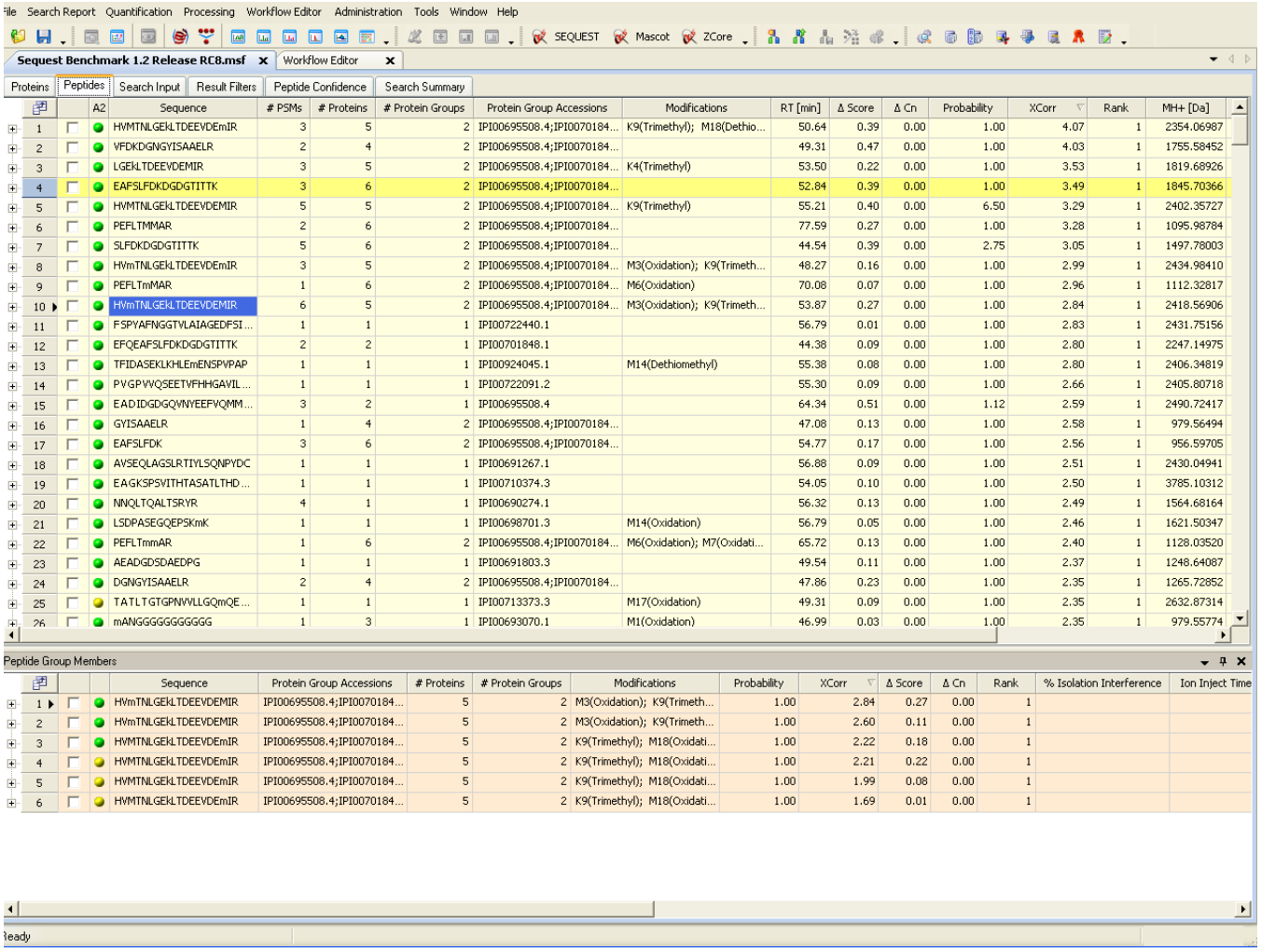
❖ To display the Peptide Group Members view

1. If the peptides are not already grouped, group them by right-clicking on the Peptides page of the MSF file and choosing **Show Peptide Groups**.
2. Click in a peptide row.

3. Choose **Search Report > Show Peptides Group Members View** or click the Show Peptides Group Members View icon, .

   The Peptides Group Members view opens at the bottom of the Proteome Discoverer window, as shown in Figure 186. In this example, the view shows the peptides that compose the HVmTNLGEklTDEEVDEMIR group.

Figure 186. Peptides Group Members view



For information on the columns shown in the Peptide Group Members view, see "Peptides Page Parameters" on page 291.

When peptide grouping is not enabled (that is, when Show Peptide Groups is not selected), the Peptide Group Members view does not display any peptides.

Calculating the Delta Score

In the Proteome Discoverer application application, each identified peptide has a ΔScore (Delta Score) column on the Peptides page. The ΔScore column displays the delta score, which is a measure of the difference between the top two scores for the peptides identified by that spectrum. The Proteome Discoverer application calculates this score as follows:

$$\Delta\text{Score} = (\text{Score}(\text{Rank } N \text{ Peptide}) - \text{Score}(\text{Rank 1 Peptide}))/\text{Score}(\text{Rank 1 Peptide})$$

The delta score is displayed only for the rank 1 peptide. For peptides with ranks greater than 1, the Delta Score column is left blank.

The Proteome Discoverer application always calculates the delta score from the score specified as the "main score" of a search engine. Table 56 shows how the main score is defined for the different search engines. The application compares the scores based on the precision with which they are displayed. This means two Ion Score values displayed for a Mascot search or two XCorr values displayed for a SEQUEST search are regarded as equal, if the displayed score value is the same.

Table 56. Main score for the search engines

Search engine	Main score	Precision (digits)
Mascot	Ion Score	0 (integer)
SEQUEST	XCorr	2
ZCore	Probability Score	2

The Proteome Discoverer application handles two special cases:

- The peptide that is ranked 1 has the same score as the peptide that is ranked 2, and both have the same sequence. In this case, the two peptides differ by their modification state. The search engine used cannot score them differently, so the ordering of these two peptides as rank 1 and rank 2 is only by chance. You can apply the calculation described in the first paragraph of this section to situations where the peptides with ranks 1 to n have the same score and the same sequence. The peptides with ranks 1 to n all display the same delta score, which is the normalized difference of the score of the rank $n + 1$ peptide.
- If no peptide is ranked 2, the peptide with rank 1 displays a delta score of 1.0. The same is true when the peptides with ranks 1 to n have the same score and the same sequence, but there is no peptide with rank $n + 1$. In this case, all peptides with rank 1 to n display a delta score of 1.0.

Using the Search Input Page

The Search Input page, shown in Figure 187, displays all the precursor ions that have been fragmented in the part of the spectrum specified by the Spectrum Selector node in the workflow. It displays detailed information about the spectra (mass peak lists) that were processed during the search process. If you used the grouping parameters, this report lists the grouped spectra. You can also check to see how many identified peptides were found and which proteins the peptide belongs to. You can also select spectra to create an Xcalibur exclusion mass list, which is explained in "Exporting Exclusion Mass Lists to Xcalibur" on page 600.

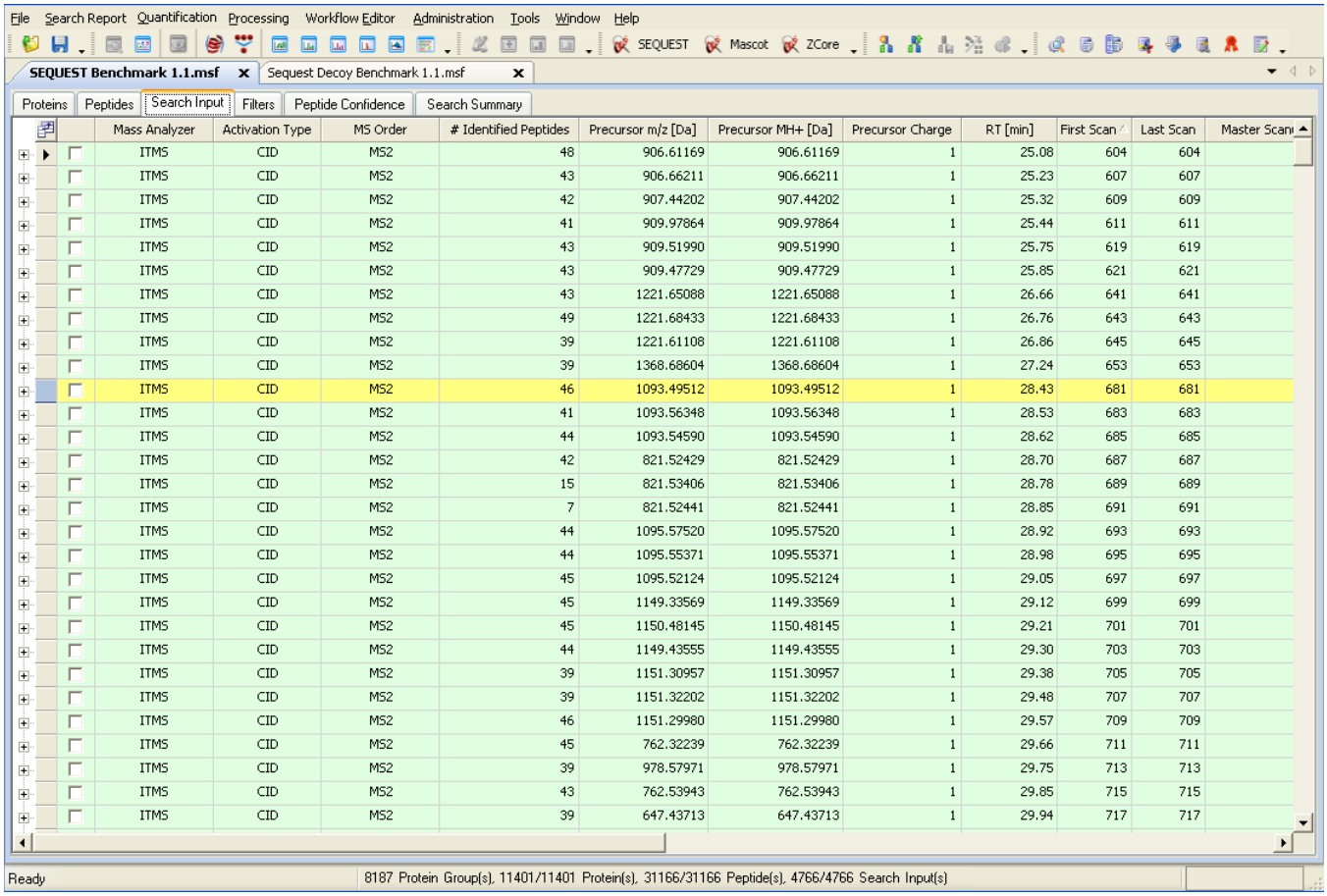
You can access the Search Input page from an open results report.

❖ To display the Search Input page

- In an open MSF file, click the **Search Input** tab.

The Search Input page of your search report appears. To select the columns to display, see "Selecting the Columns to Display" on page 223.

Figure 187. Search Input page



Mass Analyzer	Activation Type	MS Order	# Identified Peptides	Precursor m/z [Da]	Precursor MH+ [Da]	Precursor Charge	RT [min]	First Scan	Last Scan	Master Scan
ITMS	CID	MS2	48	906.61169	906.61169	1	25.08	604	604	
ITMS	CID	MS2	43	906.66211	906.66211	1	25.23	607	607	
ITMS	CID	MS2	42	907.44202	907.44202	1	25.32	609	609	
ITMS	CID	MS2	41	909.97864	909.97864	1	25.44	611	611	
ITMS	CID	MS2	43	909.51990	909.51990	1	25.75	619	619	
ITMS	CID	MS2	43	909.47729	909.47729	1	25.85	621	621	
ITMS	CID	MS2	43	1221.65088	1221.65088	1	26.66	641	641	
ITMS	CID	MS2	49	1221.68433	1221.68433	1	26.76	643	643	
ITMS	CID	MS2	39	1221.61108	1221.61108	1	26.86	645	645	
ITMS	CID	MS2	39	1368.68604	1368.68604	1	27.24	653	653	
ITMS	CID	MS2	46	1093.49512	1093.49512	1	28.43	681	681	
ITMS	CID	MS2	41	1093.56348	1093.56348	1	28.53	683	683	
ITMS	CID	MS2	44	1093.54590	1093.54590	1	28.62	685	685	
ITMS	CID	MS2	42	821.52429	821.52429	1	28.70	687	687	
ITMS	CID	MS2	15	821.53406	821.53406	1	28.78	689	689	
ITMS	CID	MS2	7	821.52441	821.52441	1	28.85	691	691	
ITMS	CID	MS2	44	1095.57520	1095.57520	1	28.92	693	693	
ITMS	CID	MS2	44	1095.55371	1095.55371	1	28.98	695	695	
ITMS	CID	MS2	45	1095.52124	1095.52124	1	29.05	697	697	
ITMS	CID	MS2	45	1149.33569	1149.33569	1	29.12	699	699	
ITMS	CID	MS2	45	1150.48145	1150.48145	1	29.21	701	701	
ITMS	CID	MS2	44	1149.43555	1149.43555	1	29.30	703	703	
ITMS	CID	MS2	39	1151.30957	1151.30957	1	29.38	705	705	
ITMS	CID	MS2	39	1151.32202	1151.32202	1	29.48	707	707	
ITMS	CID	MS2	46	1151.29980	1151.29980	1	29.57	709	709	
ITMS	CID	MS2	45	762.32239	762.32239	1	29.66	711	711	
ITMS	CID	MS2	39	978.57971	978.57971	1	29.75	713	713	
ITMS	CID	MS2	43	762.53943	762.53943	1	29.85	715	715	
ITMS	CID	MS2	39	647.43713	647.43713	1	29.94	717	717	

Search Input Page Parameters

Table 57 describes the parameters on the Search Input page.

Table 57. Search Input page parameters (Sheet 1 of 2)

Parameters	Description
Column Chooser ( )	Opens the Column Chooser dialog box, so you can choose which columns to display.
Column with boxes ( )	Allows you to select a peptide.
ID	Displays the internal number assigned by the Proteome Discoverer application that is associated with the scan number.
Search ID	Displays an internal identification number from the job queue.
Creator ID	Specifies the number of the node that creates the spectrum.
Sender ID	Displays the number of the last spectrum filter node used to pass the spectrum to the search node.
Mass Analyzer	Displays the mass analyzer used to create the raw data file. The Proteome Discoverer application recognizes the following mass analyzers: • ITMS (Ion Trap) • FTMS (Fourier Transform) • TOFMS (Time of Flight) • SQMS (Single Quad) • TQMS (Triple Quad) • SectorMS (Sector Field)
Activation Type	Displays the fragmentation method used to produce the product spectrum.
MS Order	Indicates the order of the MS spectrum: • 1 is one MS scan. • 2 is an MS/MS scan. • 3 is an MS^3 scan.
# Identified Peptides	Displays the total number of peptide matches found for that spectrum.

Table 57. Search Input page parameters (Sheet 2 of 2)

Parameters	Description
% Isolation Interference	Displays the percentage of interference by co-isolation within the precursor isolation window. For more information on co-isolation, see "Excluding Peptides with High Levels of Co-Isolation" on page 547.
Ion Inject time [ms]	Specifies the time in milliseconds used to accumulate ions in the mass spectrometer to reach their target value before they are scanned out. Low ion injection times are desirable because they indicate high transmission and fast cycle times. Ion injection times that slowly increase might indicate lower transmission, the need for multiple gain recalibration, or both.
Precursor m/z [Da]	Displays the mass-to-charge ratio of the precursor ion, in daltons.
Precursor MH+ [Da]	Displays the singly charged mass of the precursor ions, in daltons.
Precursor Charge	Displays the charge state of the precursor ion.
Precursor Intensity	Displays the intensity of the precursor ion.
RT [min]	Displays the retention time when the spectrum was detected, in minutes.
First Scan	Displays the number of the first scan in grouped scans used to produce the composite search input scan, if grouping is performed.
Last Scan	Displays the number of the last scan in grouped scans used to produce the composite search input scan, if grouping is performed. The last scan might not be the same as the first scan if the workflow enabled grouped scans.
Master Scan(s)	Displays the number of the scan of the precursor ion or ions.
Scan(s)	Displays the number of the MS/MS spectrum scan or scans.
Spectrum File	Displays the path and name of the raw data file or spectrum container file used to produce the search input spectrum.

Using the Result Filters Page

For information about the Result Filters page and instructions on filtering data with it, see "Filtering Data" on page 359.

Using the Peptide Confidence Page

For information about the Peptide Confidence page and instructions on calculating false discovery rates with it, see "Calculating False Discovery Rates" on page 415.

Using the Search Summary Page

The Search Summary page, shown in Figure 188, provides a text version of the parameters that you set to perform a search. It is divided into several categories containing information about the raw file, basic spectrum parameters, and search application-specific parameters. The categories represent the nodes used during the various search phases.

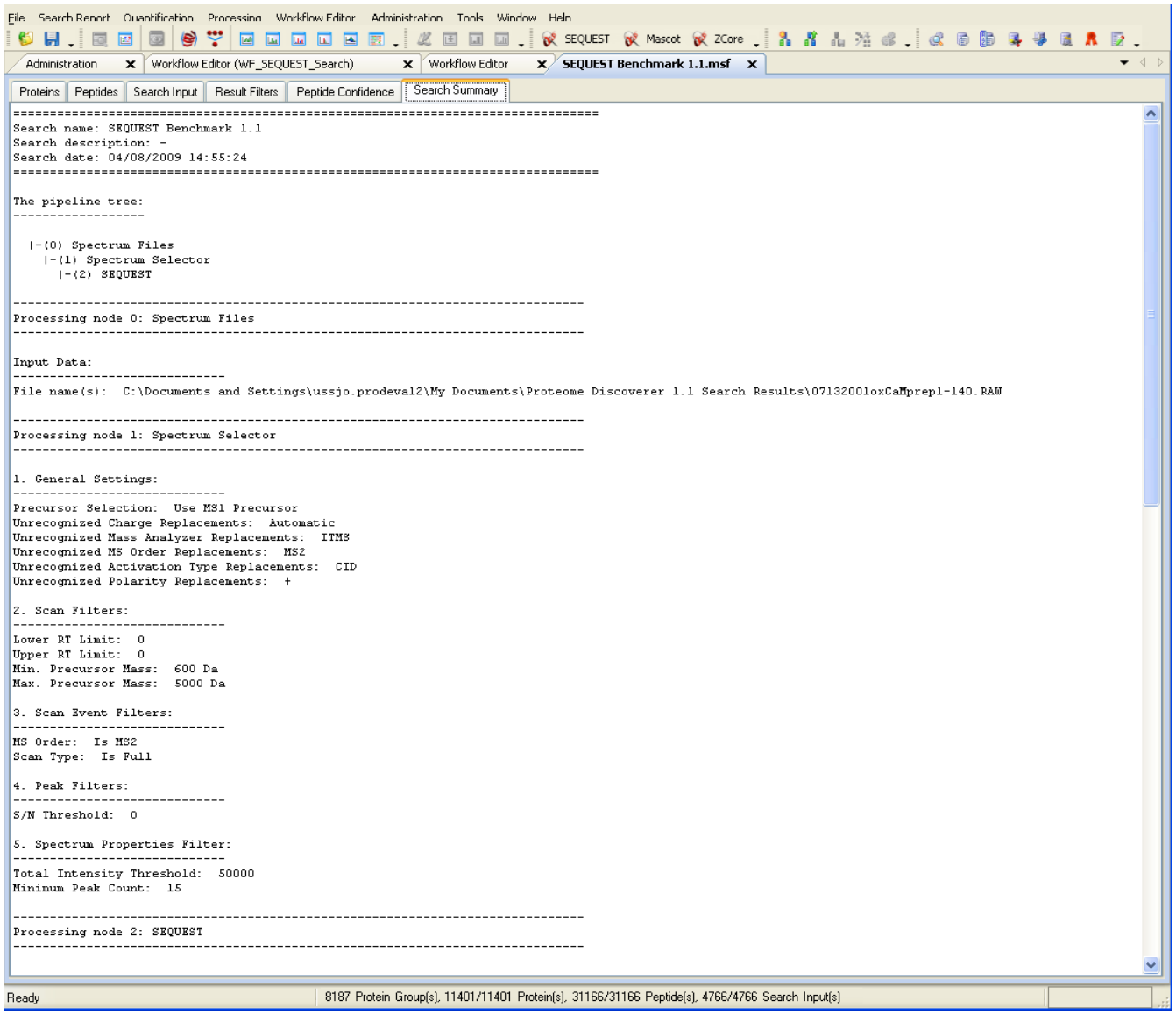
You can access the Search Summary page from an open results report.

❖ To display the Search Summary page

- In an open report, click the **Search Summary** tab.

The Search Summary page of your search report appears, as shown in Figure 188.

Figure 188. Search Summary page (partial)



Search Summary Page Parameters

Table 58 describes the parameters on the Search Summary page.

Table 58. Search Summary page parameters

Parameter	Description
Search identification section	Gives the name of the search, a description of the search, the date when the search was performed, and the order that the nodes were processed during the search.
Nodes section	Lists the nodes used in the workflow and the parameter settings for each node.
Processing details section	Lists the messages that the Proteome Discoverer application issues as it searches the raw file.

Using the Quantification Summary Page

The Quantification Summary page appears only when you open an MSF file containing quantification results. For information about the Quantification Summary page, see "Summarizing the Quantification" on page 523.

Using the Quan Spectra Page

The Quan Spectra page appears only when you open an MSF file containing quantification results from reporter ion quantification. For information about the Quan Spectra page, see "Displaying Quantification Spectra" on page 527.

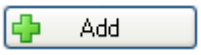
Creating a Report from Multiple Results

The Proteome Discoverer application can combine the results of multiple individual searches into one multiconsensus report. The combined report compiles information about the peptides and proteins identified in individual searches with the percent coverage.

❖ **To combine the results of multiple searches into a report**

1. In the Proteome Discoverer window, choose **File > Open Report**, click the **Open Report** icon,  , or press CTRL+O.

   The Input Files page now appears, as shown in Figure 133 on page 215.

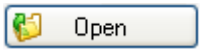
2. Click the **Add** icon, 

.

3. In the Add Analysis File(s) dialog box, browse to at least two MSF files that you want to load, select them, and click **Open**.

   The files must all reside in the same directory.

   You can use the CTRL key to select files that are not adjacent in the list in the Add Analysis File(s) dialog box.

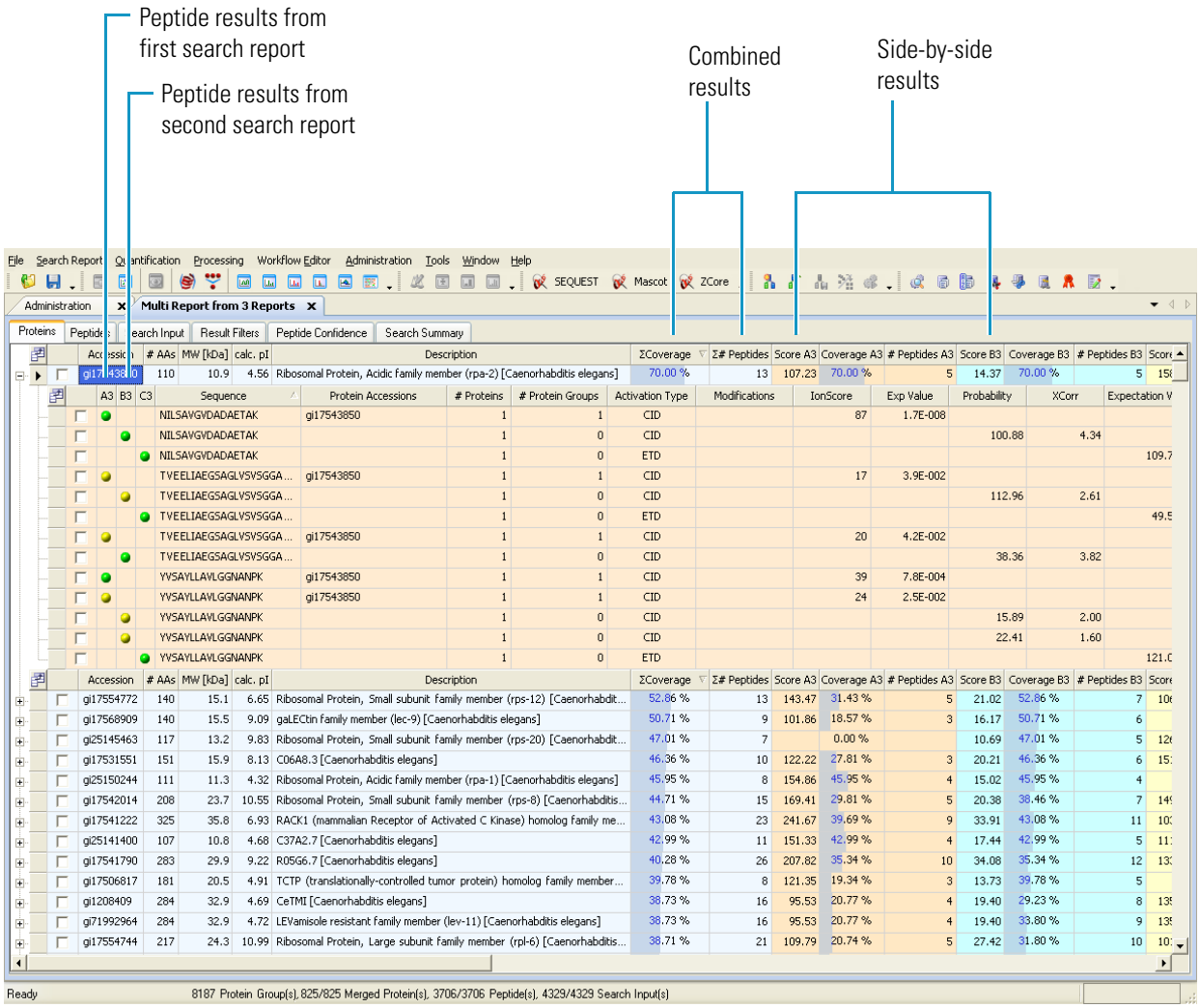
4. Click the **Open** icon, 

.

   The combined multiconsensus report appears.

Figure 189 shows a sample multiconsensus report.

Figure 189. Multiconsensus report



The combined results from the individual result files are numbered with alphanumeric labels. A number following a letter, such as A3, identifies which search node in the workflow was used to create the results file.

On the Proteins page, the individual search identification information, such as protein score, sequence coverage, and the number of identified peptides, is displayed side by side for each protein, as shown in Figure 189. The peptides identified in the individual searches are compared by sequence to align with the results in the report. When the proteins referenced in the individual searches have different accession numbers, the accession number of the first search in the result set is chosen for display.

On the Peptides page, the results shown depend on the settings chosen through its shortcut menu. For descriptions of the commands on the Peptides page shortcut menu, see "Shortcut Menus" on page 630. By default, the Proteome Discoverer application sorts by XCorr (SEQUEST), IonScore (Mascot), or Probability (ZCore). The default setting shows the top matches per peptide and search engine.

The compiled report contains results in titled and separate columns for the different searches.

When you set the display options to show the peptide matches grouped by peptide sequence, the identification information from the individual searches is displayed side by side, as shown in Figure 189. The grouped peptides do not represent actual matches found during the search but rather the unified information from all matches found for this particular peptide sequence. You can access the peptide information by examining the Proteins page for all peptides associated with a protein and the Peptides page for all peptides, including those not associated with any protein.

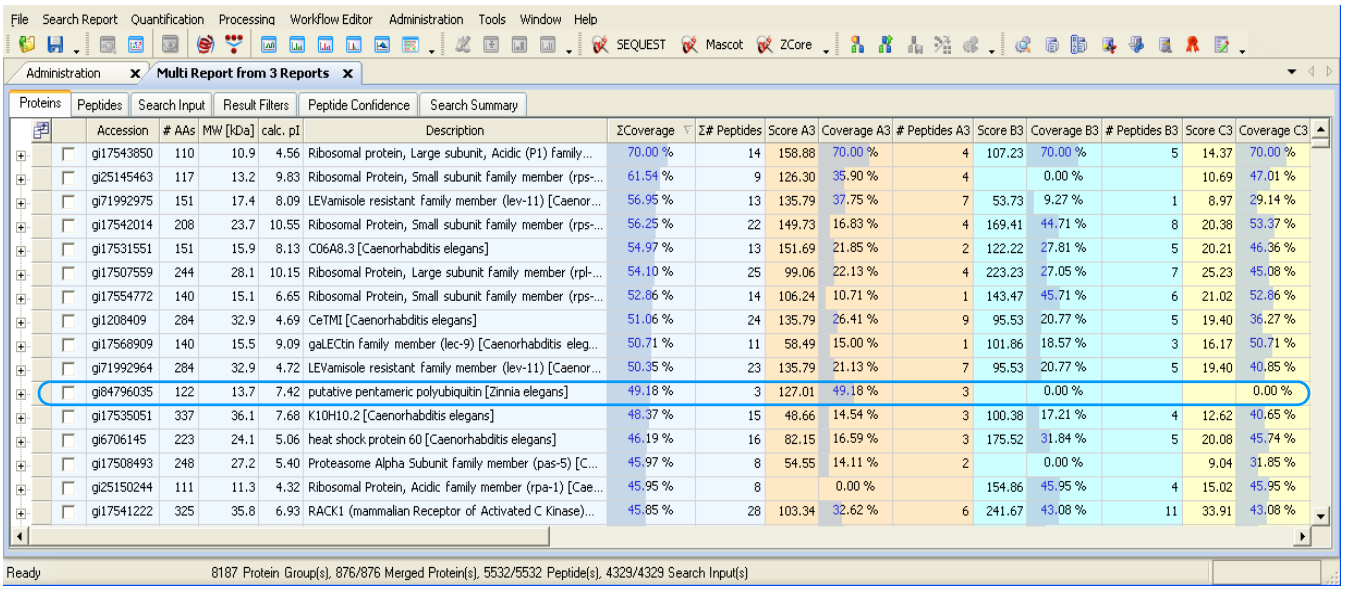
You can view the associated peptides on the Proteins page. Click **+** to the left of the protein row to access the peptide rows.

Finding Common or Unique Proteins in Multiple Searches

You can use a multiconsensus report to find common or unique proteins between two or more searches.

Figure 190 shows a multiconsensus report from three searches and the protein unique to search A.

Figure 190. Multiconsensus report showing proteins unique to search A



Accession	# AAs	MW [kDa]	calc. pI	Description	ΣCoverage	Σ# Peptides	Score A3	Coverage A3	# Peptides A3	Score B3	Coverage B3	# Peptides B3	Score C3	Coverage C3
gi17543850	110	10.9	4.56	Ribosomal protein, Large subunit, Acidic (P1) family...	70.00 %	14	158.88	70.00 %	4	107.23	70.00 %	5	14.37	70.00 %
gi25145463	117	13.2	9.83	Ribosomal Protein, Small subunit family member (rps-...	61.54 %	9	126.30	35.90 %	4		0.00 %		10.69	47.01 %
gi71992975	151	17.4	8.09	LEVamisole resistant family member (lev-11) [Caenor...	56.95 %	13	135.79	37.75 %	7	53.73	9.27 %	1	8.97	29.14 %
gi17542014	208	23.7	10.55	Ribosomal Protein, Small subunit family member (rps-...	56.25 %	22	149.73	16.83 %	4	169.41	44.71 %	8	20.38	53.37 %
gi17531551	151	15.9	8.13	C06A8.3 [Caenorhabditis elegans]	54.97 %	13	151.69	21.85 %	2	122.22	27.81 %	5	20.21	46.36 %
gi17507559	244	28.1	10.15	Ribosomal Protein, Large subunit family member (rpl-...	54.10 %	25	99.06	22.13 %	4	223.23	27.05 %	7	25.23	45.08 %
gi17554772	140	15.1	6.65	Ribosomal Protein, Small subunit family member (rps-...	52.86 %	14	106.24	10.71 %	1	143.47	45.71 %	6	21.02	52.86 %
gi1208409	284	32.9	4.69	CeTMI [Caenorhabditis elegans]	51.06 %	24	135.79	26.41 %	9	95.53	20.77 %	5	19.40	36.27 %
gi17568909	140	15.5	9.09	gaLECtin family member (lec-9) [Caenorhabditis eleg...	50.71 %	11	58.49	15.00 %	1	101.86	18.57 %	3	16.17	50.71 %
gi71992964	284	32.9	4.72	LEVamisole resistant family member (lev-11) [Caenor...	50.35 %	23	135.79	21.13 %	7	95.53	20.77 %	5	19.40	40.85 %
gi84796035	122	13.7	7.42	putative pentameric polyubiquitin [Zinnia elegans]	49.18 %	3	127.01	49.18 %	3		0.00 %			0.00 %
gi17535051	337	36.1	7.68	K10H10.2 [Caenorhabditis elegans]	48.37 %	15	48.66	14.54 %	3	100.38	17.21 %	4	12.62	40.65 %
gi6706145	223	24.1	5.06	heat shock protein 60 [Caenorhabditis elegans]	46.19 %	16	82.15	16.59 %	3	175.52	31.84 %	5	20.08	45.74 %
gi17508493	248	27.2	5.40	Proteasome Alpha Subunit family member (pas-5) [C...	45.97 %	8	54.55	14.11 %	2		0.00 %		9.04	31.85 %
gi25150244	111	11.3	4.32	Ribosomal Protein, Acidic family member (rpa-1) [Cae...	45.95 %	8		0.00 %		154.86	45.95 %	4	15.02	45.95 %
gi17541222	325	35.8	6.93	RACK1 (mammalian Receptor of Activated C Kinase)...	45.85 %	28	103.34	32.62 %	6	241.67	43.08 %	11	33.91	43.08 %

The marked protein was only identified in search A (that is, it is unique to search A), but all other proteins were identified in all three searches (that is, they are common to all three searches). What distinguishes the unique from the common proteins is the number of identified peptides for a particular protein within the different searches by the different search nodes. This value is displayed in the # Peptides column. In Figure 190, the number of identified peptides for the protein is 3 in search A but zero in searches B and C. (If no peptides were identified, the cell would be left blank to make it easier to spot in long lists of numbers). The protein was therefore identified in search A but not in searches B and C. Otherwise, at least one peptide would be identified for these proteins.

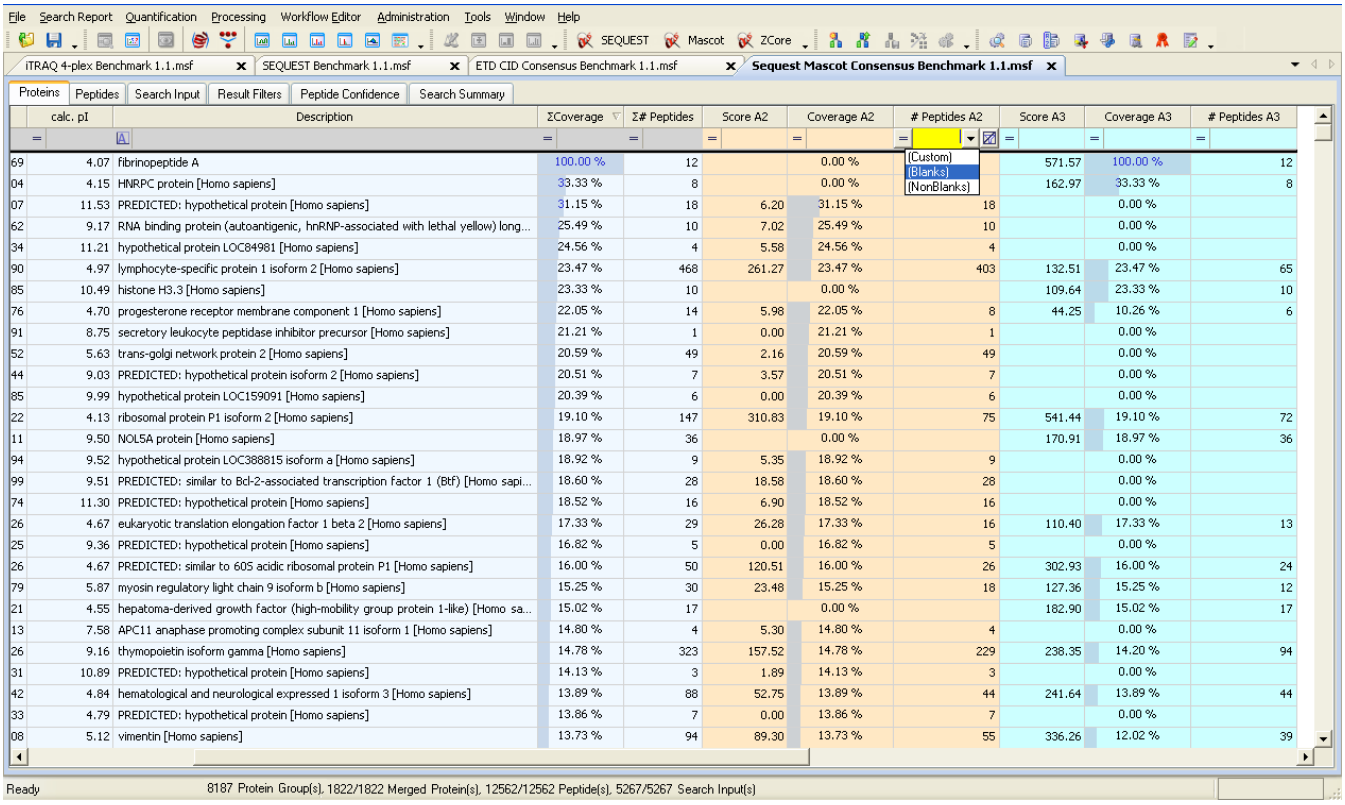
In general, if a protein is uniquely identified by one search or search node, the number of identified peptides for this protein is non-zero for this search or search node but zero in all other searches or search nodes. On the other hand, if a protein is common to two or more searches or search nodes, the number of identified peptides is non-zero for all considered searches or search nodes.

You can use the row filters of the Proteins, Peptides, or Search Input page to automate the task of creating a list of all proteins unique to one search or search node or a list of all proteins common to all searches or search nodes.

❖ To create a list of all proteins unique to one search or search node or a list of all proteins common to all searches or search nodes

1. Right-click the Proteins, Peptides, or Search Input page, and select **Enable Row Filters** from the shortcut menu.
2. Click  beneath the # Peptides A2 column header.
3. From the list, select **Blanks**, as shown in Figure 191.

Figure 191. Creating lists of unique and common proteins



	calc. pI	Description	ΣCoverage	Σ# Peptides	Score A2	Coverage A2	# Peptides A2	Score A3	Coverage A3	# Peptides A3
69	4.07	fibrinopeptide A	100.00 %	12		0.00 %		571.57	100.00 %	12
04	4.15	HNRPC protein [Homo sapiens]	33.33 %	8		0.00 %		162.97	33.33 %	8
07	11.53	PREDICTED: hypothetical protein [Homo sapiens]	31.15 %	18	6.20	31.15 %	18		0.00 %	
62	9.17	RNA binding protein (autoantigenic, hnRNP-associated with lethal yellow) long...	25.49 %	10	7.02	25.49 %	10		0.00 %	
34	11.21	hypothetical protein LOC84981 [Homo sapiens]	24.56 %	4	5.58	24.56 %	4		0.00 %	
90	4.97	lymphocyte-specific protein 1 isoform 2 [Homo sapiens]	23.47 %	468	261.27	23.47 %	403	132.51	23.47 %	65
85	10.49	histone H3.3 [Homo sapiens]	23.33 %	10		0.00 %		109.64	23.33 %	10
76	4.70	progesterone receptor membrane component 1 [Homo sapiens]	22.05 %	14	5.98	22.05 %	8	44.25	10.26 %	6
91	8.75	secretory leukocyte peptidase inhibitor precursor [Homo sapiens]	21.21 %	1	0.00	21.21 %	1		0.00 %	
52	5.63	trans-golgi network protein 2 [Homo sapiens]	20.59 %	49	2.16	20.59 %	49		0.00 %	
44	9.03	PREDICTED: hypothetical protein isoform 2 [Homo sapiens]	20.51 %	7	3.57	20.51 %	7		0.00 %	
85	9.99	hypothetical protein LOC159091 [Homo sapiens]	20.39 %	6	0.00	20.39 %	6		0.00 %	
22	4.13	ribosomal protein P1 isoform 2 [Homo sapiens]	19.10 %	147	310.83	19.10 %	75	541.44	19.10 %	72
11	9.50	NOL5A protein [Homo sapiens]	18.97 %	36		0.00 %		170.91	18.97 %	36
94	9.52	hypothetical protein LOC388815 isoform a [Homo sapiens]	18.92 %	9	5.35	18.92 %	9		0.00 %	
99	9.51	PREDICTED: similar to Bcl-2-associated transcription factor 1 (Btf) [Homo sapi...	18.60 %	28	18.58	18.60 %	28		0.00 %	
74	11.30	PREDICTED: hypothetical protein [Homo sapiens]	18.52 %	16	6.90	18.52 %	16		0.00 %	
26	4.67	eukaryotic translation elongation factor 1 beta 2 [Homo sapiens]	17.33 %	29	26.28	17.33 %	16	110.40	17.33 %	13
25	9.36	PREDICTED: hypothetical protein [Homo sapiens]	16.82 %	5	0.00	16.82 %	5		0.00 %	
26	4.67	PREDICTED: similar to 60S acidic ribosomal protein P1 [Homo sapiens]	16.00 %	50	120.51	16.00 %	26	302.93	16.00 %	24
79	5.87	myosin regulatory light chain 9 isoform b [Homo sapiens]	15.25 %	30	23.48	15.25 %	18	127.36	15.25 %	12
21	4.55	hepatoma-derived growth factor (high-mobility group protein 1-like) [Homo sa...	15.02 %	17		0.00 %		182.90	15.02 %	17
13	7.58	APC11 anaphase promoting complex subunit 11 isoform 1 [Homo sapiens]	14.80 %	4	5.30	14.80 %	4		0.00 %	
26	9.16	thymopoietin isoform gamma [Homo sapiens]	14.78 %	323	157.52	14.78 %	229	238.35	14.20 %	94
31	10.89	PREDICTED: hypothetical protein [Homo sapiens]	14.13 %	3	1.89	14.13 %	3		0.00 %	
42	4.84	hematological and neurological expressed 1 isoform 3 [Homo sapiens]	13.89 %	88	52.75	13.89 %	44	241.64	13.89 %	44
33	4.79	PREDICTED: hypothetical protein [Homo sapiens]	13.86 %	7	0.00	13.86 %	7		0.00 %	
08	5.12	vimentin [Homo sapiens]	13.73 %	94	89.30	13.73 %	55	336.26	12.02 %	39

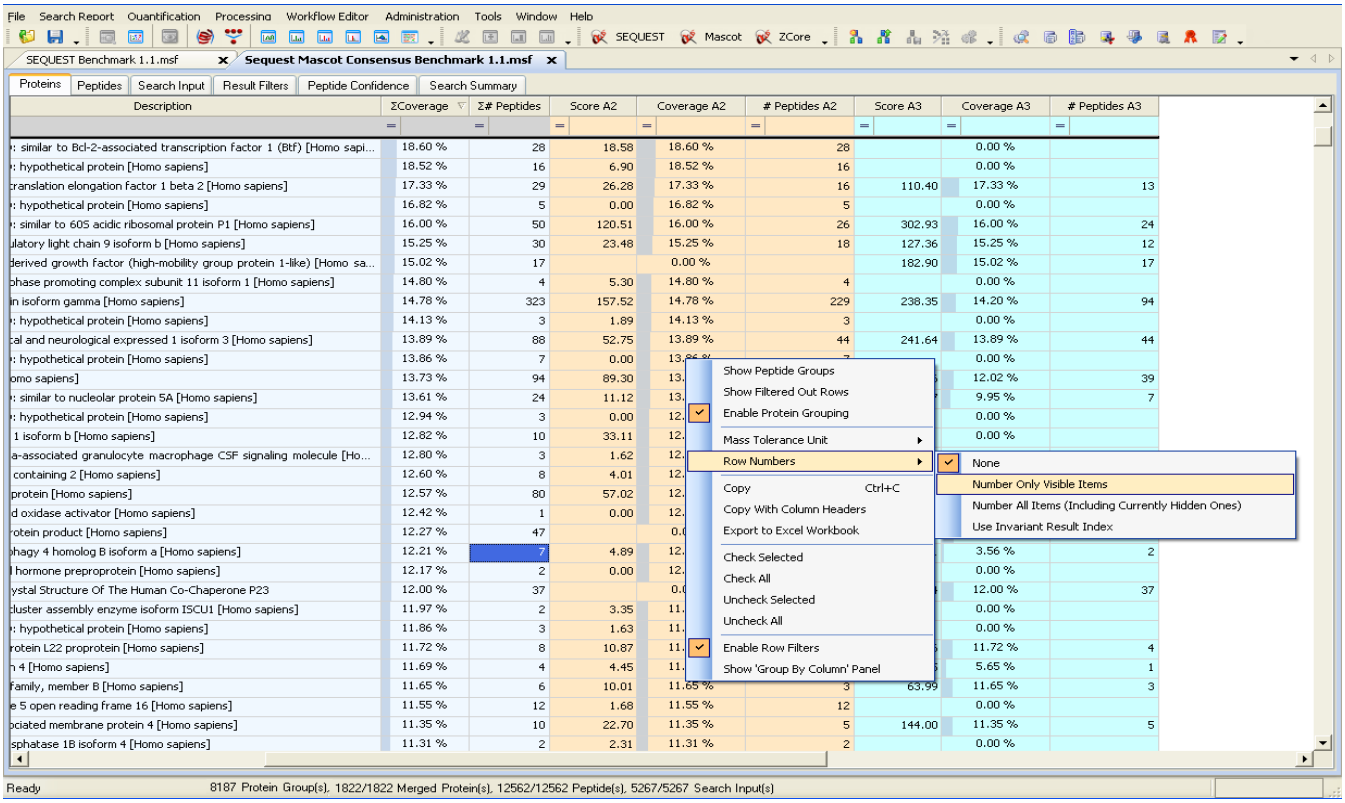
4. Click ▼ beneath the # Peptides A3 column header.
5. From the list, select **NonBlanks**.

   These steps create a list of all proteins showing that the number of identified peptides is non-zero in search A and zero in search B. This is the same as the list of proteins uniquely identified in search A.

   To automatically create a list of all proteins identified in both search A and search B, set row filters to NonBlank for both search A and search B. You can generalize these steps to multiconsensus reports from more than two searches or search nodes.

6. If you are interested in the number of proteins unique to one search or search node or common to all searches or search nodes, display the row numbers by right-clicking and choosing **Row Numbers > Number Only Visible Items** from the shortcut menu, as shown in Figure 192.

Figure 192. Displaying the number of unique or common proteins



When you load a multiconsensus report from several individual result files with quantification, you must specify whether the result files to load are the result of searching technical or biological replicates or the result of a comparison. To indicate that the result files are replicates, select the **Treat Quan Results as Replicates** option on the Input Files page when you choose File > Open Report.

Applying Filters Specific to Different Searches in Multiconsensus Reports

You cannot apply filters specific to different searches in multiconsensus reports, but you can apply filters that are specific to the different search engines used (SEQUEST, Mascot, and ZCore). The available score-based filters automatically change the list of scores that you can select for filtering, depending on the search engines used for the loaded reports.

For Mascot search report, the following peptide filters are available:

- Mascot Significance Threshold
- Peptide Score

See Table 62 for information about these filters.

For a SEQUEST search report, the following peptide filters are available:

- Peptide Score
- FScore versus Charge State

See Table 61 for information about these filters.

For a ZCore search report, the following peptide filters are available:

- Peptide Score

See Table 63 for information about this filter.

For a joint Mascot/SEQUEST search report, the following filters are available:

- Mascot Significance Threshold
- Peptide Score
- Score versus Charge State

See Table 61 and Table 62 for information about these filters.

You can combine any of these filters in a final set of filters. You can also use the same filter multiple times. For example, you can use the following filters:

- Peptide Score filter with Mascot(IonScore) > 30
- Peptide Score filter with SEQUEST(XCorr) > 20

In this case, the Mascot score filters affect only Mascot results, and the SEQUEST filters affect only SEQUEST results.

Filtering Data

The single or multiconsensus MSF report displays a list of matching peptides and proteins identified by the search engine that you specify. This chapter explains how to sort and filter the data from your Proteome Discoverer results report.

Contents

- Result Filters Page
- Filtering Data
- Grouping Proteins
- Grouping Peptides
- Calculating False Discovery Rates

Result Filters Page

On the Result Filters page, shown in Figure 193, you can select the proteins and peptides to filter out of the search results. Refining your search results in this way can make your analysis quicker. By using filters, you can sort and filter your results by charge state, modifications, or even peptide probability. You can also create and apply more than one filter to your search results.

In addition to the Result Filters page, you can filter the data while opening your MSF file by setting filters on the Result Filters page that appears when you choose File > Open Report (see Figure 135 on page 218). These filters are identical to the filters on the Result Filters page for an already opened MSF file, except that you can only set protein filters on the Result Filters page for an already opened MSF file.

Protein scores give some indication of the relevance of a protein. They are calculated from a list of peptides identified for a particular protein and can be expected to change as soon as the peptides are removed by the application of result filters. The Proteome Discoverer application recalculates the protein scores after you apply peptide filters or change the score thresholds on the Peptide Confidence page. For information on how the application calculates protein scores, see Table 39.

Displaying the Result Filters Page

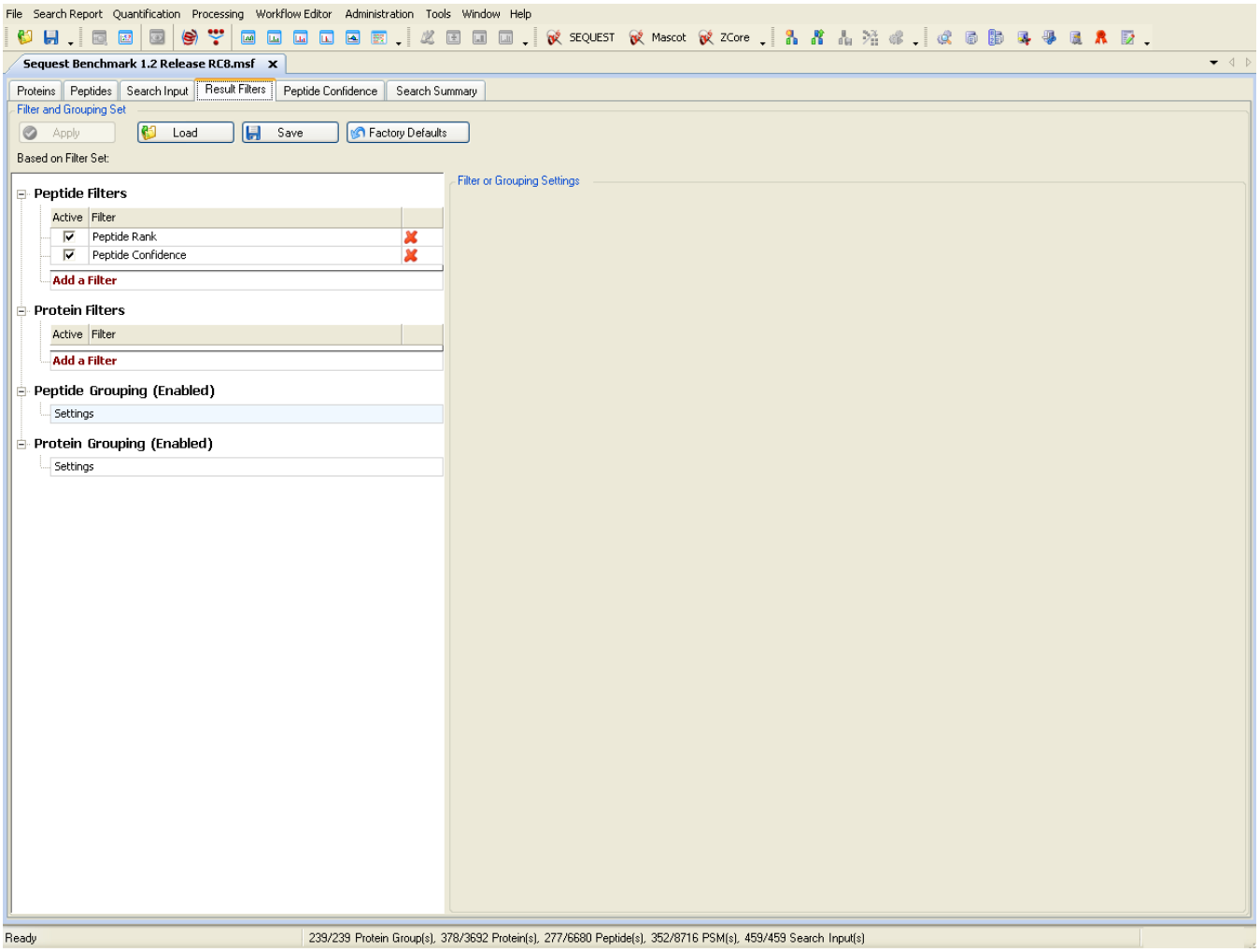
To filter the number of proteins and peptides visible on the Proteins and Peptides pages in an MSF file that is already open, use the Result Filters page, shown in Figure 193. For information about filtering MSF files while opening them, see "Opening MSF Files" on page 215.

❖ To display the Result Filters page in an open MSF file

- In an open report, click the **Result Filters** tab.

  The Result Filters page of your results report appears, as shown in Figure 193.

Figure 193. Result Filters page



Result Filters Page Parameters

Table 59 describes the parameters on the Result Filters page. These parameters differ by type of search engine.

Table 59. Result Filters page parameters (Sheet 1 of 3)

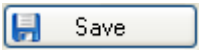
Parameter	Description
Filter and Grouping Set	Specifies the settings for peptide and protein filters and protein grouping.
	Applies the selected filters or protein grouping setting to the data in the search results report. This option becomes available only after you modify or move a filter to the filter specification box or change the protein grouping setting.
	Loads the selected filter set.
	Opens the Save Filter Set dialog box so that you can save the selected filters as a filter set.
	Restores the filter settings to those in effect when you first installed the Proteome Discoverer application.
Peptide Filters	Specifies the filters for peptides. • For SEQUEST peptide filters, see Table 61. • For Mascot peptide filters, see Table 62. • For ZCore peptide filters, see Table 63.
Active	Adds the filter to the filter set. You must also click Apply to apply the filter set to your current results. When you clear the Active check box, the filter is temporarily removed from the filter set. By default, a newly added filter is selected and active.
Filter	Displays the name of the filter.
Add a Filter	Adds a new peptide filter.
Protein Filters	Specifies the filters for proteins. See Table 60.
Active	Adds the filter to the filter set. You must also click Apply to apply the filter set to your current results. When you clear the Active check box, the filter is temporarily removed from the filter set. By default, a newly added filter is selected and active.
Filter	Displays the name of the filter.

Table 59. Result Filters page parameters (Sheet 2 of 3)

Parameter	Description
Add a Filter	Adds a new protein filter.
Peptide Grouping (Enabled)	Specifies the grouping for peptides.
Settings	Sets filter and peptide grouping parameters.
Filter or Grouping Settings	Specifies how to group peptides: • Show Peptide Groups: Determines whether peptides are shown grouped on the Peptides page. • Group Peptides by: Specifies how to group peptides: – Mass and Sequence: Groups peptides by both mass and sequence. – Sequence: Groups peptides by sequence.
Protein Grouping (Enabled)	Specifies the grouping for proteins.

Table 59. Result Filters page parameters (Sheet 3 of 3)

Parameter	Description
Settings	Sets filter and protein grouping parameters.
Filter or Grouping Settings	Specifies how to group proteins: • Enable Protein Grouping: Determines whether homologous proteins are grouped. This option is selected by default. • Consider Leucine and Isoleucine as Equal: Determines whether leucine and isoleucine are treated as equal amino acids. This option is selected by default. • Consider Only PSMs with Confidence at Least: Determines the type of PSMs that the Proteome Discoverer application considers for inclusion in protein grouping: • Low: Considers low-confidence PSMs for inclusion in protein grouping. • (Default) Medium: Considers medium-confidence PSMs for inclusion in protein grouping. • High: Considers high-confidence PSMs for inclusion in protein grouping. • Consider Only PSMs with Delta Cn Better Than: Considers only PSMs with values lower than or equal to the specified value for inclusion in the protein grouping process. The default Delta Cn value is 0.15. • Apply Strict Maximum Parsimony Principle: Determines whether the Proteome Discoverer application removes all protein groups that are not necessary to explain the found peptides. This option is clear by default.

Protein Filters

Table 60 lists the protein filters available in results reports.

Table 60. Protein filters (Sheet 1 of 2)

Filter type	Description
Differentiable Proteins	Specifies whether to display proteins that contain at least one unique peptide but also share one or more peptides with another, or more than one, protein. • Include Distinct Proteins: Displays proteins that contain at least one unique peptide.
Distinct Proteins	Displays only the matches containing the set number of unique peptides (the range is from 0 to 100).
Peptides Per Protein	Specifies the minimum number of peptides per protein and displays all proteins with a number higher than the specified number. For example, if you specify a minimum of two peptides per protein, the Proteome Discoverer application retains all proteins having two or more peptides. The filter only counts peptides with different sequences. It only counts peptides with the same sequence but different modifications as one. It also only counts peptides with isobaric sequence changes (for example, isoleucine against leucine) as one. • Minimal Number of Peptides: Specifies the minimum number of peptides per protein. • Count Only Rank 1 Peptides: Includes only peptides with a rank of 1. • Count peptide Only in Top Scored Proteins: Includes the peptides only in the top-scoring proteins.
Protein Description Contains	Retains all proteins containing a certain word or phrase in their descriptions in the Description column. • Annotation Substring: Specifies the word or phrase in the protein description that causes the protein to be retained. • Ignore case: Determines whether the search for the designated word or phrase is case-sensitive.

Table 60. Protein filters (Sheet 2 of 2)

Filter type	Description
Protein Score	Filters out proteins that have a score below the specified threshold. • Search Node: Specifies the node that identified the proteins to be filtered. • Score Threshold: Specifies the threshold score below which proteins will be filtered out. The minimum value is 0.0, and there is no maximum value.
Unique Proteins	Filters out proteins that are found in all files or by all search nodes. This filter is only available when you open either a multiconsensus report containing data from at least two files or a file containing data performed by two search nodes.

Peptide Filters

The peptide filters available depend on the search engine used.

- Peptide Filters in SEQUEST Reports
- Peptide Filters in Mascot Reports
- Peptide Filters in ZCore Reports

Peptide Filters in SEQUEST Reports

Table 61 lists the peptide filters available in reports generated by the SEQUEST search engine.

Table 61. SEQUEST peptide filters (Sheet 1 of 5)

Filter type	Description
Activation Type	Filters out all peptides that were not fragmented by the specified activation type. Activation Type: • Activation Types: Specifies the activation type. You can choose the following activation types: – CID (Collision-Induced Dissociation) – MPD (Multi-Photon Dissociation – ECD (Electron Capture Dissociation) – PQD (Pulsed Q Collision-Induced Dissociation) – ETD (Electron Transfer Dissociation) – HCD (High-Energy Collision Dissociation) – Any For more information about these activation types, see "Fragmentation Methods" on page 8.
Charge State	Filters out all peptides that do not have the specified charge state. Charge State: • Lowest Charge State: Filters out peptides with a charge lower than this state. The minimum value is 0, and there is no maximum value. The default is 0. • Highest Charge State: Filters out peptides with a charge higher than this state. The minimum value is zero, and there is no maximum value if you select a value of 0. The default is 0.

Table 61. SEQUEST peptide filters (Sheet 2 of 5)

Filter type	Description
Has Modifications	Retains all peptides that have modifications. Has Modifications: • Peptides That Are: – Unmodified: Peptides are not modified by post-translational modifications (PTMs), static modifications, or dynamic modifications – (Default) Modified: Peptides are modified by post-translational modifications (PTMs), static modifications, or dynamic modifications.
Peptide Confidence	Filters out all peptides that do not have the selected peptide confidence level. Peptide Confidence: • Minimum Confidence: Specifies the minimum peptide confidence level: – (Default) High: Displays peptides of high confidence, indicated by green circles on the Peptides page of the MSF report. – Medium: Displays peptides of medium confidence, indicated by yellow circles on the Peptides page of the MSF report. – Low: Displays peptides of low confidence, indicated by red circles on the Peptides page of the MSF report.
Peptide Delta Cn	Specifies the Delta Cn threshold, which filters out all PSMs with a Delta Cn larger than the specified value. For more information on this value, see "Filtering Peptides by Delta Cn Value" on page 385. Peptide Delta Cn: • Maximum Delta Cn: Specifies a Delta Cn threshold and filters out all PSMs with a Delta Cn larger than this value. The minimum value is 0.0, and the maximum value is 1.0.

Table 61. SEQUEST peptide filters (Sheet 3 of 5)

Filter type	Description
Peptide Length	Filters out peptides with a length less than the lowest specified value and a length greater than the highest specified value. Peptide Length: • Lowest Peptide Length: Filters out peptides with a length less than the specified value. • Highest Peptide Length: Filters out peptides with a length greater than the specified value.
Peptide Mass Deviation	Filters out all peptides with a mass deviation higher than the specified mass deviation. Peptide Mass Deviation: • Mass Deviation: Specifies the mass deviation, in daltons, parts per million, or milli-mass units.
Peptide Rank	Filters out peptides with a rank higher than the rank that you specify in the Maximum Peptide Rank box. For detailed information about the Peptide Rank filter, see "Filtering Peptides by Rank" on page 382. Peptide Rank: • Maximum Rank: Specifies the maximum peptide rank. The minimum value is 1, and there is no maximum value.

Table 61. SEQUEST peptide filters (Sheet 4 of 5)

Filter type	Description
Peptide Score	Filters out all peptides with scores lower than the score specified in the Score Threshold box. Peptide Score: • Peptide Score: Filters out peptides with scores less than the score entered in the Score Threshold box for the designated score type. You can choose from these score types: – SequestNode (ProbabilityScore): Filters out peptides with scores less than the Score Threshold score in the Probability column on the Peptides page. – SequestNode (SpScore): Filters out peptides with scores less than the Score Threshold score in the SpScore column on the Peptides page. – (Default) SequestNode (XCorr): Filters out peptides with scores less than the Score Threshold score in the XCorr column on the Peptides page. • Score Threshold: Specifies the peptide score below which to filter out all peptides. You can specify a value between 0 and 1000, inclusive. The default is 10.
Peptides Containing Any of the Specified Amino Acids	Retains all peptides containing any of the specified amino acids. Peptides Containing Any of the Specified Amino Acids: • Amino Acids (One Letter Codes): Specifies the letter of the amino acid contained in the peptides to be retained.

Table 61. SEQUEST peptide filters (Sheet 5 of 5)

Filter type	Description
Precursor Mass	Filters out all peptides having a precursor mass that does not fall in the specified precursor mass range. Precursor Mass: • Lowest Precursor Mass: Specifies the beginning of the precursor mass range, in daltons. The default is 600 Da. • Highest Precursor Mass: Specifies the end of the precursor mass range, in daltons. The default is 3500 Da.
Score versus Charge State	Filters out peptides having a charge state of *x* and a score value lower than the specified score. Score versus Charge State: • Score to Which the Filter Is Applied: Filters out peptides with scores less than the score entered in the Score Threshold box for the designated score type. You can choose from these score types: – SequestNode (ProbabilityScore): Filters out peptides with scores less than the Score Threshold score in the Probability column on the Peptides page. – SequestNode (SpScore): Filters out peptides with scores less than the Score Threshold score in the SpScore column on the Peptides page. – (Default) SequestNode (XCorr): Filters out peptides with scores less than the Score Threshold score in the XCorr column on the Peptides page. • Minimal Score for Charge State = *x*: Specifies the score of the peptide for the *x* charge state. Values can range from 0.0 to 1000.0. Default values vary. • Minimal Score for Charge State > *x*: Specifies the score of the peptide greater than the *x* charge state. Values can range from 0.0 to 1000.0. The default is 3.4.

Peptide Filters in Mascot Reports

Table 62 lists the peptide filters available in reports generated by the Mascot search engine.

Table 62. Mascot peptide filters (Sheet 1 of 4)

Filter type	Description
Activation Type	Filters out all peptides that were not fragmented by the specified activation type. Activation Type: • Activation Types: Specifies the activation type. You can choose the following activation types: – CID (Collision-Induced Dissociation) – MPD (Multi-Photon Dissociation) – ECD (Electron Capture Dissociation) – PQD (Pulsed Q Collision-Induced Dissociation) – ETD (Electron Transfer Dissociation) – HCD (High-Energy Collision Dissociation) – Any For more information about these activation types, see "Fragmentation Methods" on page 8.
Charge State	Filters out all peptides that do not have the specified Peptide Charge State Filter. Charge State: • Lowest Charge State: Filters out peptides with a charge lower than this state. The minimum value is 0, and there is no maximum value. The default is 0. • Highest Charge State: Filters out peptides with a charge higher than this state. The minimum value is zero, and there is no maximum value if you select a value of 0. The default is 0.

Table 62. Mascot peptide filters (Sheet 2 of 4)

Filter type	Description
Has Modifications	Retains all peptides that have modifications. Has Modifications: • Peptides That Are: – Unmodified: Peptides are not modified by post-translational modifications (PTMs), static modifications, or dynamic modifications – (Default) Modified: Peptides are modified by post-translational modifications (PTMs), static modifications, or dynamic modifications.
Mascot Significance Threshold	Filters out all peptides with a score lower than the specified significance threshold in the IonScore column. Mascot Significance Threshold: • Significance Threshold: Specifies the threshold below which peptides are filtered out. Use a value between 0 and 1. The default is 0.05.
Peptide Confidence	Filters out all peptides that do not have the specified peptide confidence. This parameter also sets the confidence. Peptide Confidence • Minimum Confidence: Specifies the minimum peptide confidence level: – (Default) High: Displays peptides of high confidence, indicated by green circles on the Peptides page of the MSF report. – Medium: Displays peptides of medium confidence, indicated by yellow circles on the Peptides page of the MSF report. – Low: Displays peptides of low confidence, indicated by red circles on the Peptides page of the MSF report.

Table 62. Mascot peptide filters (Sheet 3 of 4)

Filter type	Description
Peptide Delta Cn	Specifies the Delta Cn threshold, which filters out all PSMs with a Delta Cn larger than the specified value. For more information on this value, see "Filtering Peptides by Delta Cn Value" on page 385. Peptide Delta Cn: • Maximum Delta Cn: Specifies a Delta Cn threshold and filters out all PSMs with a Delta Cn larger than this value. The minimum value is 0.0, and the maximum value is 1.0.
Peptide Length	Filters out peptides with a length less than the lowest specified value and a length greater than the highest specified value. Peptide Length: • Lowest Peptide Length: Filters out peptides with a length less than the specified value. • Highest Peptide Length: Filters out peptides with a length greater than the specified value.
Peptide Mass Deviation	Filters out all peptides with a mass deviation higher than the specified mass deviation. Peptide Mass Deviation: • Mass Deviation: Specifies the mass deviation, in daltons, parts per million, or milli-mass units.
Peptide Rank	Filters out peptides with a rank higher than the rank that you specify in the Maximum Peptide Rank box. It can also filter out peptides with a rank higher than the specified rank but with a score identical to that of the last included peptide. For detailed information about the Peptide Rank filter, see "Filtering Peptides by Rank" on page 382. Peptide Rank: • Maximum Rank: Specifies the maximum peptide rank. The minimum value is 1, and there is no maximum value.

Table 62. Mascot peptide filters (Sheet 4 of 4)

Filter type	Description
Peptide Score	Filters out all peptides with scores lower than the score specified in the Score Threshold box. Peptide Score: • Peptide Score: Filters out peptides with scores less than the designated score type. – Mascot (ExpectationValue): Filters out peptides with scores less than the value in the Expectation Value column on the Peptides page. – ((Default) Mascot (IonScore): Filters out peptides with scores less than the Score Threshold score in the IonScore column on the Peptides page. • Score Threshold: Specifies the peptide score below which to filter out all peptides. You can specify a value between 0 and 1000, inclusive. The default is 10.
Peptides Containing Any of the Specified Amino Acids	Retains all peptides containing the specified amino acids. Peptides Containing Any of the Specified Amino Acids: • Amino Acids (One Letter Codes): Specifies the letter of the amino acid contained in the peptides to be retained.
Precursor Mass	Filters out all peptides having a precursor mass that does not fall in the specified precursor mass range. Precursor Mass: • Lowest Precursor Mass: Specifies the beginning of the precursor mass range, in daltons. The default is 600 Da. • Highest Precursor Mass: Specifies the end of the precursor mass range, in daltons. The default is 3500 Da.

Peptide Filters in ZCore Reports

Table 63 lists the peptide filters available in reports generated by the ZCore search engine.

Table 63. ZCore peptide filters (Sheet 1 of 4)

Filter type	Description
Activation Type	Filters out all peptides that were not fragmented by the specified activation type. Activation Type: • Activation Types: Specifies the activation type. You can choose the following activation types: • CID (Collision-Induced Dissociation) • MPD (Multi-Photon Dissociation) • ECD (Electron Capture Dissociation) • PQD (Pulsed Q Collision-Induced Dissociation) • ETD (Electron Transfer Dissociation) • HCD (High-Energy Collision Dissociation) • Any For more information about these activation types, see "Fragmentation Methods" on page 8.
Charge State	Filters out all peptides that do not have the specified charge state. Charge State: • Lowest Charge State: Filters out peptides with a charge lower than this state. The minimum value is 0, and there is no maximum value. • Highest Charge State: Filters out peptides with a charge higher than this state. The minimum value is zero, and there is no maximum value if you select a value of 0.

Table 63. ZCore peptide filters (Sheet 2 of 4)

Filter type	Description
Has Modifications	Retains all peptides that have modifications. Has Modifications: • Peptides That Are: – Unmodified: Peptides are not modified by post-translational modifications (PTMs), static modifications, or dynamic modifications – (Default) Modified: Peptides are modified by post-translational modifications (PTMs), static modifications, or dynamic modifications.
Peptide Confidence	Filters out all peptides that do not have the specified peptide confidence. This parameter also sets the confidence. Peptide Confidence: • Minimum Confidence: Specifies the minimum peptide confidence level: – (Default) High: Selects peptides of high confidence, indicated by green circles on the Peptides page of the MSF report. – Low: Selects peptides of low confidence, indicated by red circles on the Peptides page of the MSF report. – Middle: Selects peptides of medium confidence, indicated by yellow circles on the Peptides page of the MSF report.
Peptide Delta Cn	Specifies the Delta Cn threshold, which filters out all PSMs with a Delta Cn larger than the specified value. For more information on this value, see "Filtering Peptides by Delta Cn Value" on page 385. Peptide Delta Cn: • Maximum Delta Cn: Specifies a Delta Cn threshold and filters out all PSMs with a Delta Cn larger than this value. The minimum value is 0.0, and the maximum value is 1.0.

Table 63. ZCore peptide filters (Sheet 3 of 4)

Filter type	Description
Peptide Length	Filters out peptides with a length less than the lowest specified value and a length greater than the highest specified value. Peptide Length: • Lowest Peptide Length: Filters out peptides with a length less than the specified value. • Highest Peptide Length: Filters out peptides with a length greater than the specified value.
Peptide Mass Deviation	Filters out all peptides with a mass deviation higher than the specified mass deviation. Peptide Mass Deviation: • Mass Deviation: Specifies the mass deviation, in daltons, parts per million, or milli-mass units.
Peptide Rank	Filters out peptides with a rank higher than the rank that you specify in the Maximum Peptide Rank box. For detailed information about the Peptide Rank filter, see "Filtering Peptides by Rank" on page 382. Peptide Rank: • Maximum Rank: Specifies the maximum peptide rank. The minimum value is 1, and there is no maximum value.

Table 63. ZCore peptide filters (Sheet 4 of 4)

Filter type	Description
Peptide Score	Filters out all peptides with scores lower than the score specified in the Score Threshold box. Peptide Score: • Peptide Score: Filters out peptides with scores less than the score entered in the Score Threshold box for the designated score type. You can choose from these score types: – (Default) EquestNode (ProbabilityScore): Filters out peptides with scores less than the Score Threshold score in the Probability column on the Peptides page. – EquestNode (ExpectationScore): Filters out peptides with scores higher than the Score Threshold score in the Expectation Value column on the Peptides page. • Score Threshold: Specifies the peptide score below which to filter out all peptides. You can specify a value between 0 and 1000, inclusive. The default is 10.
Peptides Containing Any of the Specified Amino Acids	Retains all peptides containing the specified amino acids. Peptides Containing Any of the Specified Amino Acids: • Amino Acids (One Letter Codes): Specifies the letter of the amino acid contained in the peptides to be retained.
Precursor Mass	Filters out all peptides having a precursor mass that does not fall in the specified precursor mass range. Precursor Mass: • Lowest Precursor Mass: Specifies the beginning of the precursor mass range, in daltons. The default is 600 Da. • Highest Precursor Mass: Specifies the end of the precursor mass range, in daltons. The default is 3500 Da.

Filtering Data

You can use Proteome Discoverer application filters to selectively hide and sort the visible results of the matched search results. You have two methods of filtering your search results data:

- Results filters on the Result Filters page exclude peptides and proteins from the results on the Proteins and Peptides pages. Applying these filters to filter out peptides, does the following:
  - Changes the number of identified peptides and the percentages shown in the Coverage column of the Proteins page.
  - Affects the numbers of filtered peptides and proteins versus the total number of peptides and proteins displayed in the Result Items Per File area at the bottom of the Input Files/Result Filters page.
  - Affects the quantification results of proteins.

  For information about filtering with the Result Filters page, see "Filtering Results with the Filters on the Result Filters Page" on page 379.
- Row filters on the shortcut menu of the Proteins, Peptides, and Search Input pages display filters only. Use these filters with the filters on the Result Filters page to narrow your search results even further. When you display the filtered-out rows, the affected lines for both filters are seen as unavailable rows. Excluding peptides by setting row filters does not change the number of identified peptides and the percentage coverage values of the proteins. For information about filtering with row filters, see "Filtering Results with Row Filters" on page 393.

If you save your report, you can save the filters that you set in the Result Filters page with your results report. You cannot save the filters that you set with the row filters with your results report. The row filters only work on the visible rows in the report. However, you can save the row filters in a saved layout. For information about saving layouts, see "Creating and Applying Custom Layouts" on page 233.

Filtering Results with the Filters on the Result Filters Page

The following procedures describe how to filter your results using the results filters on the Result Filters page.

- Filtering Your Search Results with Peptide Filters
- Filtering Peptides by Rank
- Filtering Peptides by Delta Cn Value
- Filtering Your Search Results with Protein Filters

Filtering Your Search Results with Peptide Filters

Follow this procedure to apply peptide filters to your search results.

❖ To filter your search results with peptide filters

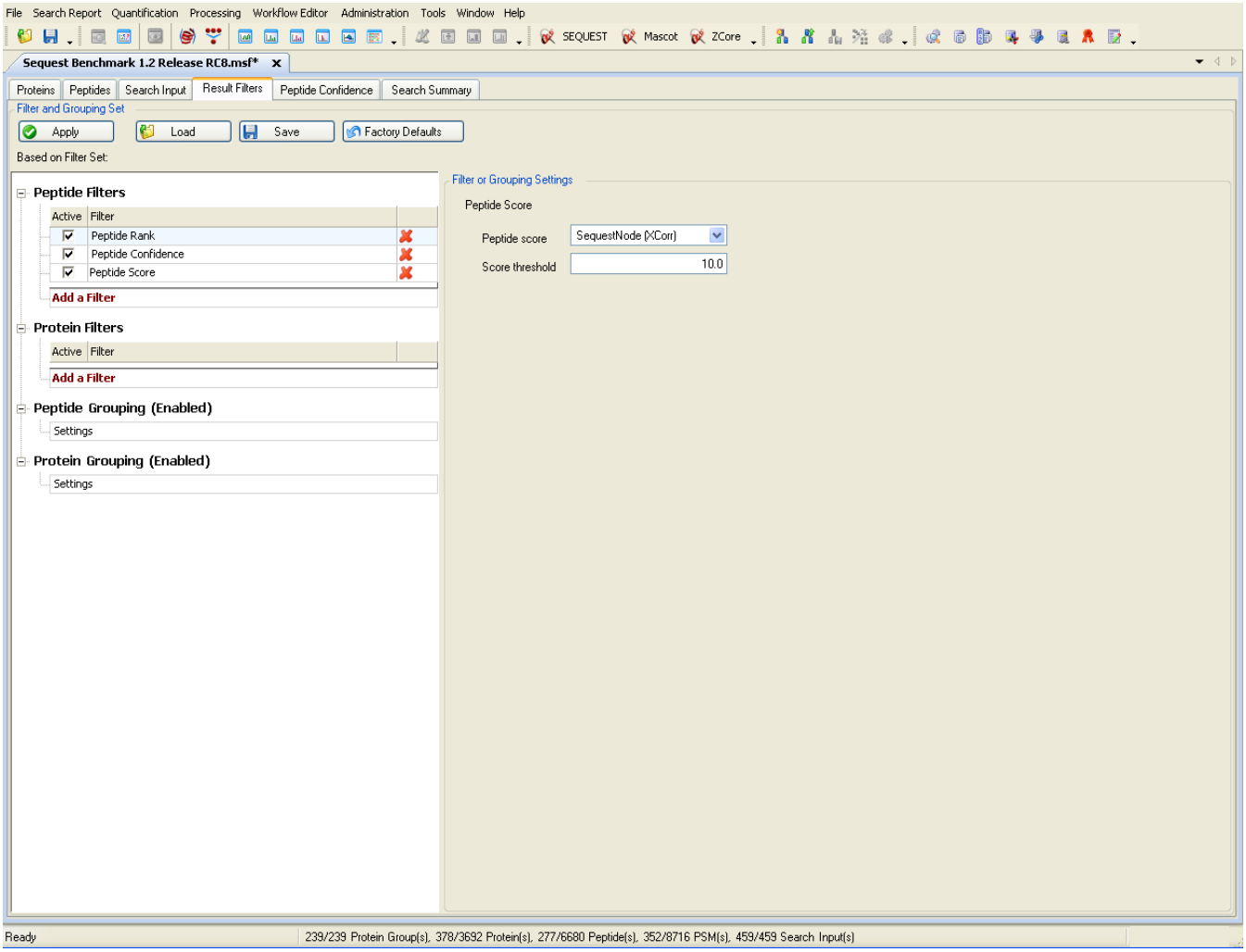
1. Open your search results. See "Opening MSF Files" on page 215.
2. Click the **Result Filters** tab, which is shown in Figure 193 on page 360.
3. Click **Add a Filter** in the Peptide Filters area.

   A list of filters now appears. For a description of the filters available, see Table 61 for SEQUEST filters, Table 62 for Mascot filters, and Table 62 for ZCore filters.

   The Peptide Rank and Peptide Confidence filters are selected by default.
4. Select the filter to apply from the list of filters.

   Options pertaining to the selected filter now appear in the Filter or Grouping Settings area, as shown for the Peptide Score filter in Figure 194. Table 61, Table 62, and Table 62 describe these options.

Figure 194. Peptide filter options



5. Set the options pertaining to the selected filter in the Filter or Grouping Settings area. For example, in Figure 194, you can set the Peptide Score option and the Score Threshold option.
6. If it is not already selected, select the check box in the Active column (it is selected by default).
7. To remove a filter before you apply it, click .
8. To update the search results, click Apply in the Filter and Grouping Set area.

Note The Proteome Discoverer application might take some time to display the filtered data.

Filtering Peptides by Rank

From the acquired MS/MS spectra, search engines like SEQUEST or Mascot create a list of possible peptides whose masses match the measured mass of the precursor ions of the MS/MS spectrum and whose fragmentation patterns match the peaks detected in the MS/MS spectrum. The better the match, the better the score of every peptide candidate considered. The Proteome Discoverer application ranks all considered peptide candidates by their scores and reports a user-specified number of peptide candidates per spectrum. The default is usually 10. The rank of a peptide is its position in the reported list of identified peptide candidates per spectrum that is ordered from better to worse scores. Peptides with a top ranking (for example, 1 or 2) are more likely to be the correct peptide than peptides with a lower ranking (for example, less than 2).

The Proteome Discoverer application does not store the peptide rank in the results file but calculates it after loading the results file. Only loaded peptides affect the peptide rank. The Proteome Discoverer application loads peptides that pass all other peptide filters before applying the Peptide Rank filter. It rejects those peptides that do not pass the Peptide Rank filter.

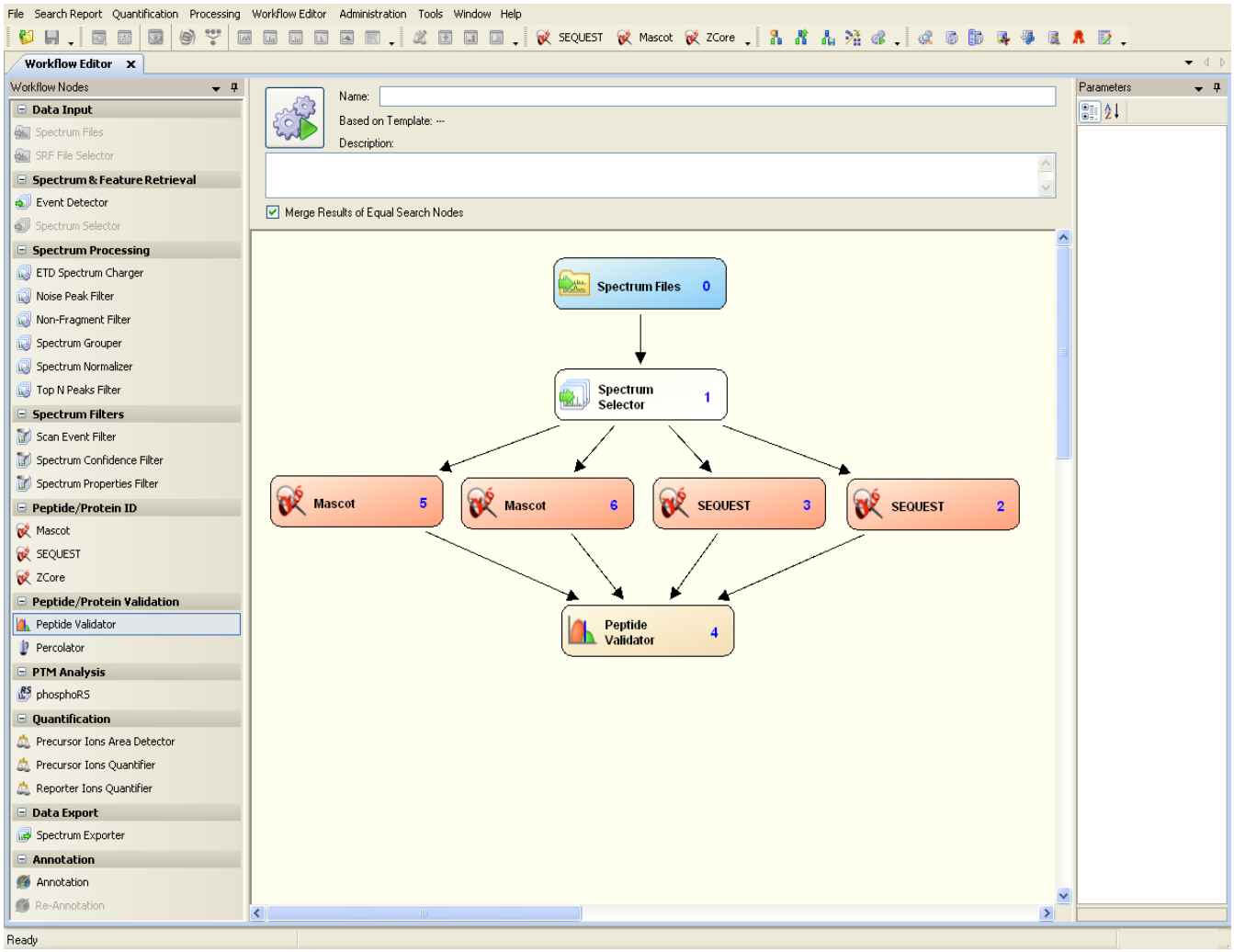
You can use the Peptide Rank filter to filter out peptides with a rank higher than the maximum rank that you specify with the Maximum Peptide Rank option.

Calculating Peptide Rank

The Merge Results of Equal Search Nodes option in the Workflow Editor determines whether peptides and proteins identified by the same type of search engine are merged together. If you select this option, the Proteome Discoverer application ranks the peptides identified by the same search engine together. Only one peptide can have rank 1 for each spectrum and search engine. If you do not select the Merge Results of Equal Search Nodes option, the Proteome Discoverer application ranks peptides identified by one search engine independently from the peptides identified by another search engine. Therefore, there can be multiple peptides having rank 1 for each spectrum.

For example, consider the workflow with two SEQUEST nodes and two MASCOT nodes shown in Figure 195.

Figure 195. Workflow with two Mascot nodes and two SEQUEST nodes



The search engines find the peptides shown in Table 64 for spectrum 10:

Table 64. Peptides found for spectrum 10

SEQUEST (2)	SEQUEST (3)	Mascot (4)	Mascot (5)
Peptide 2.1 (XCorr = 20)	Peptide 3.1 (XCorr = 12)	Peptide 4.1 (IonScore = 33)	Peptide 5.1 (IonScore = 34)
Peptide 2.2 (XCorr = 8)	Peptide 3.1 (XCorr =12)		

If you selected the Merge Results of Equal Search Nodes option, peptides 4.1 and 4.2, which Mascot identified, are ranked together. Peptides 2.1, 2.2, 3.1, and 3.2, which SEQUEST identified, are ranked together.

If you did not select the Merge Results of Equal Search Nodes option, peptides 4.1 and 4.2, which Mascot identified, are ranked independently. SEQUEST-identified peptides 2.1 and 2.2 are ranked together, and peptides 3.1 and 3.2 are ranked together.

To calculate the rank, the Proteome Discoverer application sorts all peptides belonging together by their main score. For SEQUEST, the main score is XCorr. For Mascot, the main score is IonScore. For ZCore, the main score is Probability. Peptides with the same main score have the same rank.

For example, if you selected the Merge Results of Equal Search Nodes option, the Proteome Discoverer application ranks the peptides shown in Table 64 as follows:

SEQUEST:

- Peptide 2.1 (XCorr = 20): Rank 1
- Peptide 3.1 (XCorr = 12): Rank 2
- Peptide 3.2 (XCorr = 12): Rank 2
- Peptide 2.2 (XCorr = 8): Rank 4

Mascot:

- Peptide 5.1 (IonScore =n 34): Rank 1
- Peptide 4.1 (IonScore = 33): Rank 2

If you did not select the Merge Results of Equal Search Nodes option, the Proteome Discoverer application ranks the peptides shown in Table 64 as follows:

SEQUEST (2):

- Peptide 2.1 (XCorr = 20): Rank 1
- Peptide 2.2 (XCorr = 8): Rank 2

SEQUEST (3):

- Peptide 3.1 (XCorr = 12): Rank 1
- Peptide 3.2 (XCorr = n12): Rank 1

Mascot (4):

- Peptide 4.1 (IonScore = 33): Rank 1

Mascot (5):

- Peptide 5.1 (IonScore = 34): Rank 1

Recalculating Peptide Rank

The Proteome Discoverer application does not consider filtered-out peptides in calculating peptide ranks. Filtered-out peptides have a rank of infinite. If you apply filters to an open MSF report, the application recalculates the peptide ranks. It also recalculates the delta score values each time that the peptide ranks change.

Peptide Rank Filter

If you use the Peptide Rank filter when you open a report, the Proteome Discoverer application reads the peptides twice. In the first step, it collects identifications and the main scores of all peptides passing the peptide filters except the Peptide Rank filter. Then it calculates the ranks for these peptides and loads all peptides having a higher rank than the maximum allowed rank. It loads the remaining peptides in the second step.

If you apply the Peptide Rank filter to an open report, the application filters out those peptides that do not pass the peptide filters except the Peptide Rank filter. It calculates the ranks for the remaining peptides. Finally, it applies the Peptide Rank filter.

❖ To filter peptides by rank

1. Open the MSF file. See "Opening MSF Files" on page 215.
2. Click the **Result Filters** tab.
3. Select **Peptide Rank** in the Peptide Filters area of the Result Filters page, if it is not already selected.

   The Maximum Peptide Rank option appears in the middle of the Result Filters page.
4. (Optional) In the Maximum Peptide Rank box, set the maximum rank that a peptide must have to avoid being filtered out.

   The minimum value is 1, and there is no maximum value. The default value is 1.

Filtering Peptides by Delta Cn Value

Search engines often provide multiple possible matching peptides as explanations for the same spectrum. Most of the time you can clearly distinguish the top-scoring match from the other PSMs, but sometimes, especially in the presence of dynamic modifications, the best-scoring matches of the same spectrum have very similar scores. In this case, you can filter the results to select the best-scoring PSMs and the matches that have very similar scores by using the Delta Cn peptide filter.

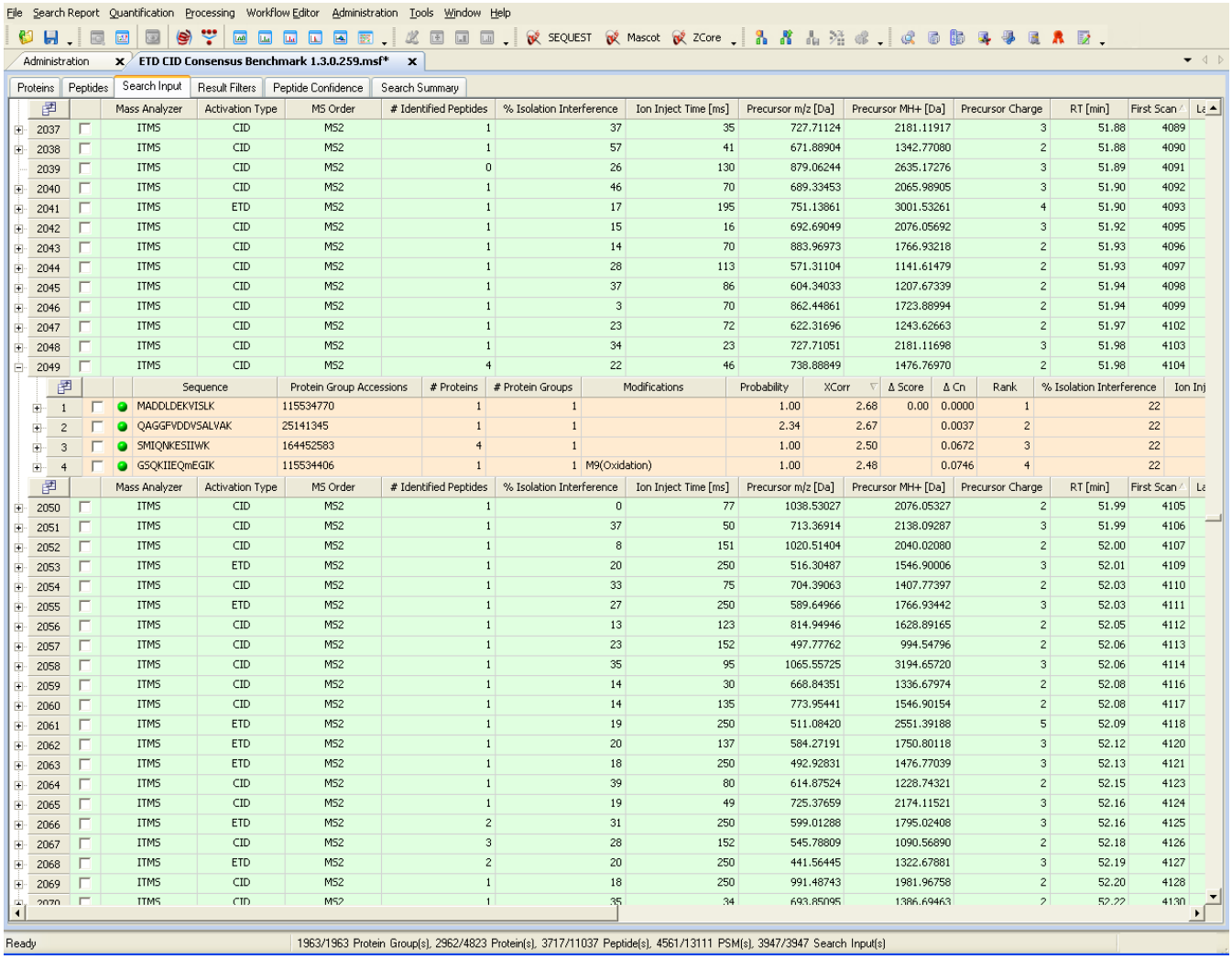
The Delta Cn value displays the normalized score difference between the currently selected PSM and the highest-scoring PSM for that spectrum:

$$\Delta Cn(\text{rank i}) = \frac{\text{score}_{\text{rank 1}} - \text{score}_{\text{rank1}}}{\text{score}_{\text{rank 1}}}$$

The Delta Cn peptide filter filters out all PSMs with a Delta Cn score larger than the specified value.

On the Peptides page, the Δ Cn column displays the Delta Cn values. For example, Figure 196 shows how the score of a peptide ranked 2 compares to other multiple high-confidence peptides from the same spectrum.

Figure 196. Delta Cn scores for multiple high-confidence peptides from the same spectrum



❖ To filter peptides by Delta Cn value

1. Open the MSF file. See "Opening MSF Files" on page 215.
2. Click the **Result Filters** tab.
3. Select **Delta Cn** in the Peptide Filters area of the Result Filters page.

   The Peptide Delta Cn option appears in the middle of the Result Filters page.

4. (Optional) In the Maximum Delta Cn box, specify a Delta Cn threshold that will filter out all PSMs with a Delta Cn larger than this value.

   The minimum value is 0.0, and the maximum value is 1.0.

Filtering Results by the Original Rank Assigned by the Search Engine

If you apply PSM-level result filters, the Proteome Discoverer application dynamically recalculates the displayed ranks, delta scores, and Delta Cn values. However, you can also view the original rank assigned by the search engine for all PSMs and peptide groups by displaying the Search Engine Rank column on the Peptides page. In addition, you can filter by this rank.

You might find this feature helpful in a case like this; assume that you know that your raw data has a true mass accuracy below 5 ppm. If you search this data with a precursor tolerance of 5 ppm and validate it by calculating false discovery rates, you obtain false positive matches within this mass deviation tolerance. You could find some of these incorrect matches if you searched the data with a larger precursor tolerance, such as 50 ppm. This step increases the chance of replacing incorrect matches with a mass deviation below 5 ppm by incorrect matches with a higher mass deviation. When you review the results, you can set a mass deviation filter of more than 5 ppm to remove all matches that have a mass deviation outside the true mass accuracy. You can now find many of the remaining incorrect matches. They have a Search Engine Rank worse than rank 1, because they were initially replaced by incorrect matches with a larger mass deviation.

Filtering Your Search Results with Protein Filters

Follow this procedure to apply protein filters to your search results.

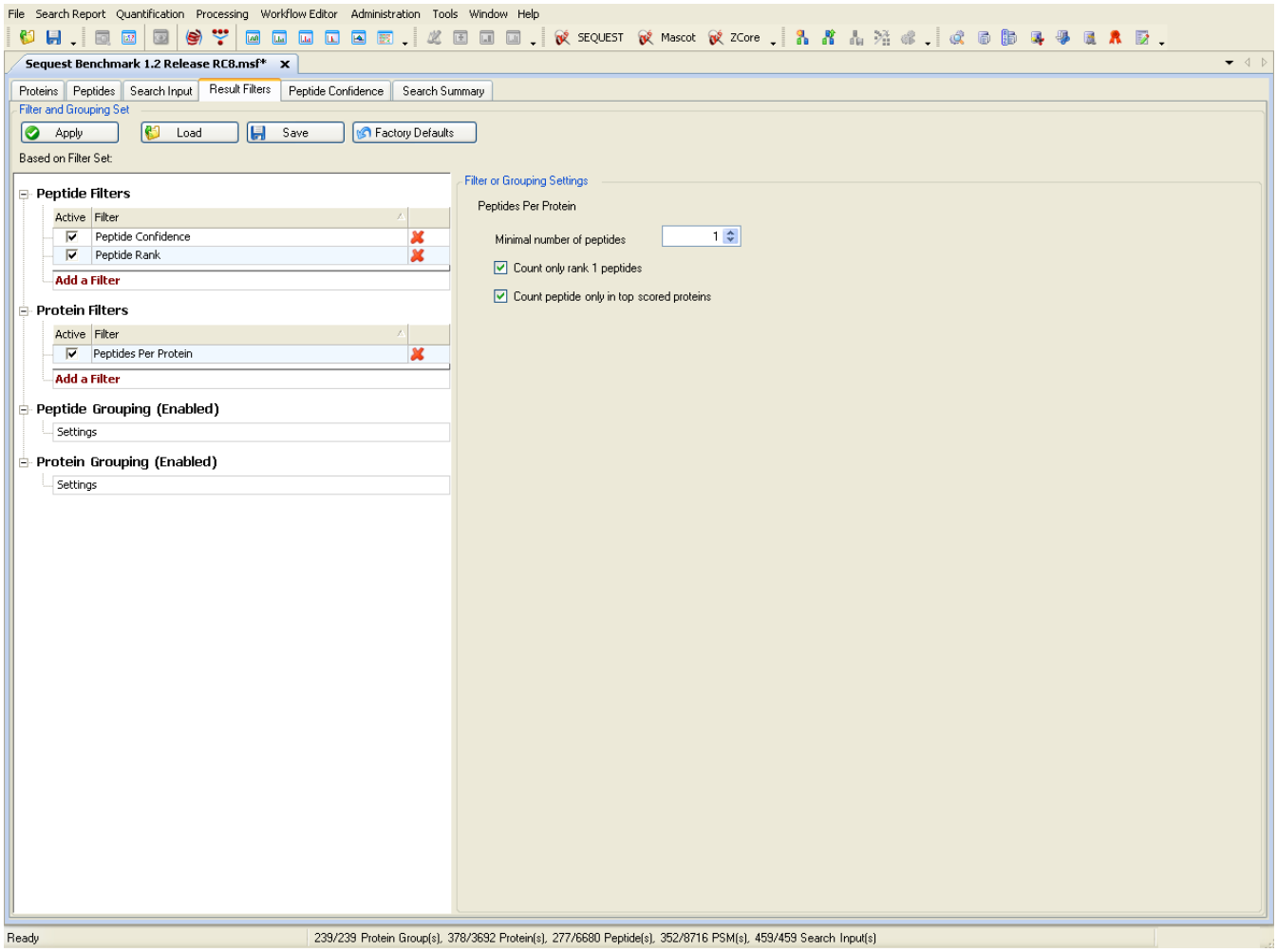
❖ To filter your search results with protein filters

1. Open your search results. See "Opening MSF Files" on page 215.
2. Click the **Result Filters** tab, which is shown in Figure 193 on page 360.
3. Click **Add a Filter** in the Protein Filters area.

   A list of filters appears. For a description of the filters available, see Table 60.
4. Select the filter to apply from the list of filters.

   Settings pertaining to the selected filter appear in the Filter or Grouping Settings area. See for a description of the available settings, see Table 60.

Figure 197. Protein filter options



5. Set the options pertaining to the selected filter in the Filter or Grouping Settings area. For example, in Figure 197, you can set the Minimal Number of Peptides, Count Only Rank 1 Peptides, and Count Peptide Only in Top Scored Proteins options.
6. If it is not already selected, select the check box in the Active column. (The check box is selected by default.)
7. To remove a filter before you apply it, click .
8. To update the search results, click Apply in the Filter and Grouping Set area.

Note The Proteome Discoverer application might take some time to display the filtered data.

Using Filter Sets

You can save your selected filter settings as a group for future use. You can also save your protein and peptide grouping settings as a set. You can make this set the default or assign it a name. These sets are saved in and loaded from external files so that you can export filter sets from one instance of the Proteome Discoverer application and import them into another instance. The filter sets have an extension of .filters.

If you want to use a filter set from one installation of the Proteome Discoverer application in another installation of the Proteome Discoverer application, you must copy the filter set from the root directory of the first installation to the root directory of the other installation.

You can create these filter setting groups on the Result Filters page that appears during report loading or on the Result Filters page that appears after the report has already been opened.

You can load a previously stored filter set. Loading a filter set replaces the currently set peptide and protein filters and the settings for the protein grouping with the filters and settings stored in the loaded filter set, unless the filters were loaded before the MSF file was opened.

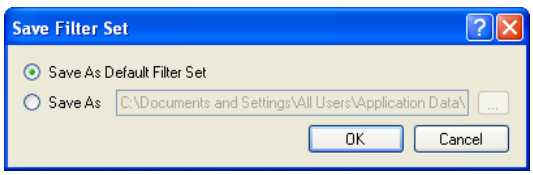
- To create and save a filter set
- To load a filter set
- To delete a filter set
- To clear the default filter set
- To restore the default filter set in effect after installing the Proteome Discoverer application

❖ To create and save a filter set

1. For filters, select the appropriate protein and peptide filters, as described in "To filter your search results with peptide filters" on page 380 and "To filter your search results with protein filters" on page 387, and click Apply.
2. In the Filter and Grouping Set area, click Save.

   The Save Filter Set dialog box appears, as shown in Figure 198.

   Figure 198. Save Filter Set dialog box



3. In the Save Filter Set dialog box, do one of the following:

   - To save the filter set or set of protein grouping settings as the default filter set, select the **Save As Default Filter Set** option.

     The Proteome Discoverer application automatically applies this filter set to the opened MSF results file.

     –or–

     To save the filter set in a file, select the **Save As**. option. Click the Browse button (**...**) and browse to the file to save it in. You can also type the name of a new file in the box next to Save As.

   - Click **OK** in the Save Filter Set dialog box.

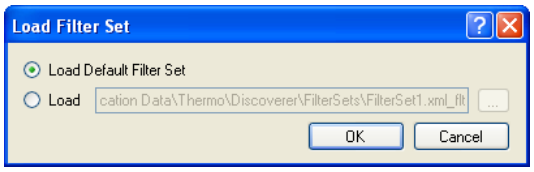
   The saved filter set appears in the list in the Filter and Grouping Set area. The default set is named "Default" in this list.

❖ To load a filter set

1. In the Filter and Grouping Set area, click Load.

   The Load Filter Set dialog box appears, as shown in Figure 199.

   Figure 199. Load Filter Set dialog box

2. In the Load Filter Set dialog box, do the following:

   a. To load the default filter set, select the **Load Default Filter Set** option.

      –or–

      To load another filter set, click the Browse button (...), and select the file containing the filter set that you want to load. You can also type the name and path of the file to load in the box next to Load.

   b. Click **OK** in the Load Filter Set dialog box.

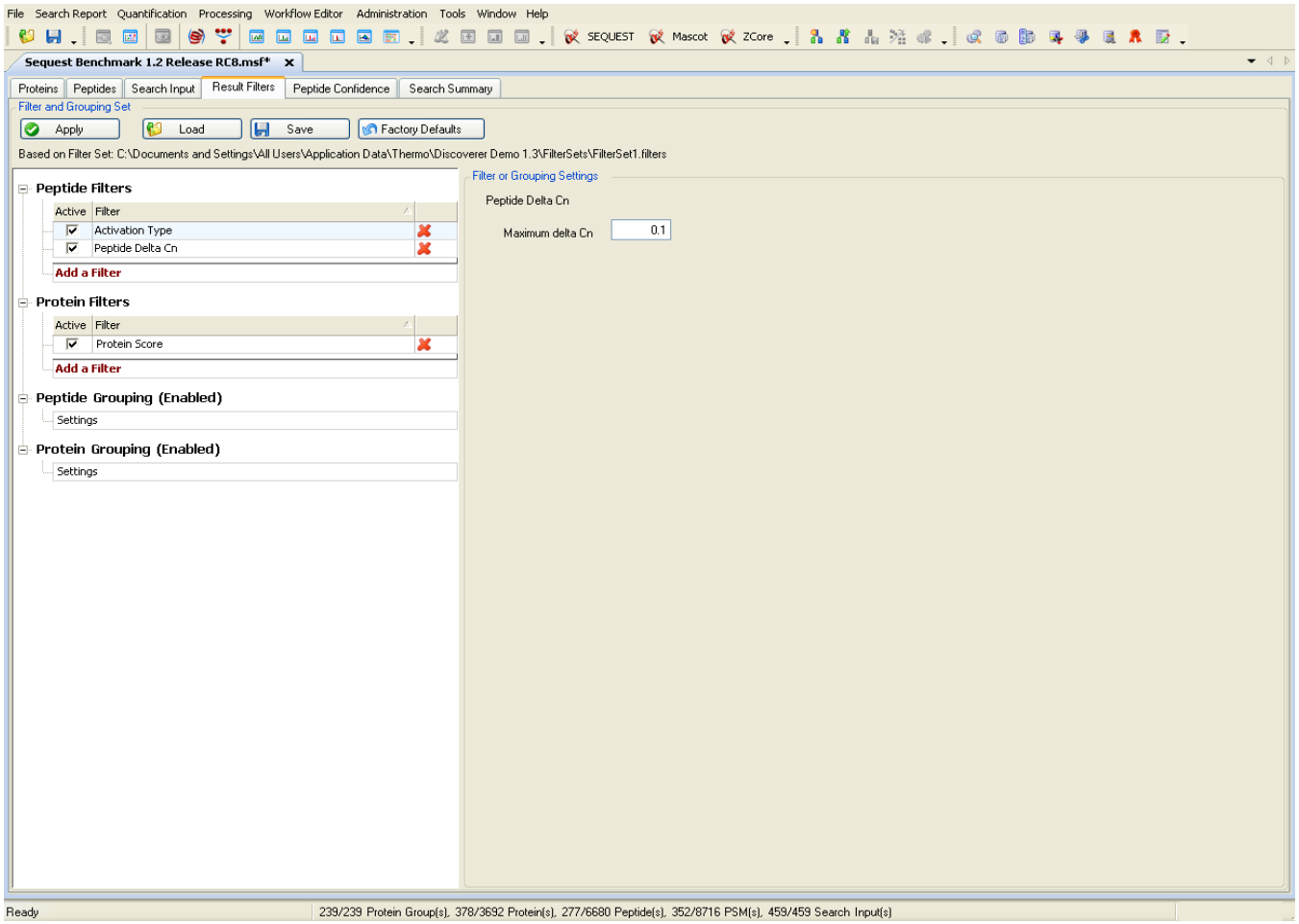
   A Loading Filter Set confirmation box appears if you have already selected other filter settings.

3. If the Loading Filter Set confirmation box appears, click **OK**.

4. If you are loading a filter set on the Results Filter page in an open MSF file, click Apply. If you are loading a filter set on the Results Filter page during report loading, the Proteome Discoverer application automatically applies the filters or sets.

The filters in the selected filter set appear in the Filter and Grouping Set area of the page, as shown in Figure 200.

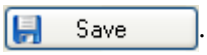
Figure 200. Loaded filter set



❖ To delete a filter set

- Click  next to the peptide filters, protein filters, or grouping sets that compose the set.

❖ To clear the default filter set

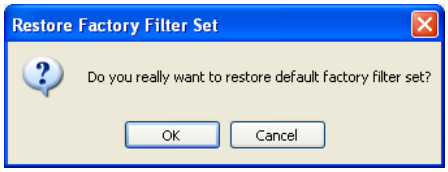
1. Remove all the peptide and protein filters from the Result Filters page by clicking  next to the peptide filters, protein filters, or grouping sets that compose the set.
2. Click  .
3. In the Save Filter Set dialog box, shown in Figure 198 on page 389, select the **Save As Default Filter Set** option.
4. Click **OK**.

❖ To restore the default filter set in effect after installing the Proteome Discoverer application

1. Click Factory Defaults.

   The confirmation box shown in Figure 201 appears.

   Figure 201. Restore Factory Filter Set confirmation box

2. Click **OK**.

   The confirmation box shown in Figure 202 appears.

   Figure 202. Loading Filter Set confirmation box

3. Click **OK**.

Removing and Deactivating Filters

You can remove or deactivate filters to alter the search results.

- To remove a filter
- To deactivate a filter

❖ To remove a filter

1. Open your search results.
2. Click the **Result Filters** tab.
3. Select the filter in the list of filters in the Peptide Filters or Protein Filters area.
4. Click .

   The filter is removed from the list of filters.

5. Click **Apply** to update the Proteins, Peptides, or Search Input page.

❖ To deactivate a filter

1. Open your search results.
2. Click the **Result Filters** tab.

3. Clear the check box in the Active column.
4. To update the Proteins, Peptides, or Search Input page, click **Apply**.

   The filter is deactivated but not removed from the Result Filters page.

Filtering Results with Row Filters

You can use row filters on the Proteins, Peptides, and Search Input pages to set up simple filter criteria that only consist of a single filter statement, such as "*number* is greater than 5," or "*text* contains kinase."

- To filter your search results using row filters
- To set a precursor mass filter by using the row filter menu
- To clear all filter conditions set by the row filter menu
- To clear an individual filter set by the row filter menu

❖ To filter your search results using row filters

1. Open your search results.
2. Select the **Proteins**, **Peptides**, or **Search Input** page.
3. Ungroup the peptides by right-clicking and choosing **Show Peptide Groups**.
4. Right-click to access the shortcut menu and choose **Enable Row Filters**.

   A filter row appears beneath the column header that contains the icons shown in Figure 203. You can select an operator, enter the filter value, clear the currently set filter, or open the Enter filter criteria for *header_name* dialog box for more complex transactions.

Figure 203. Row filter icons

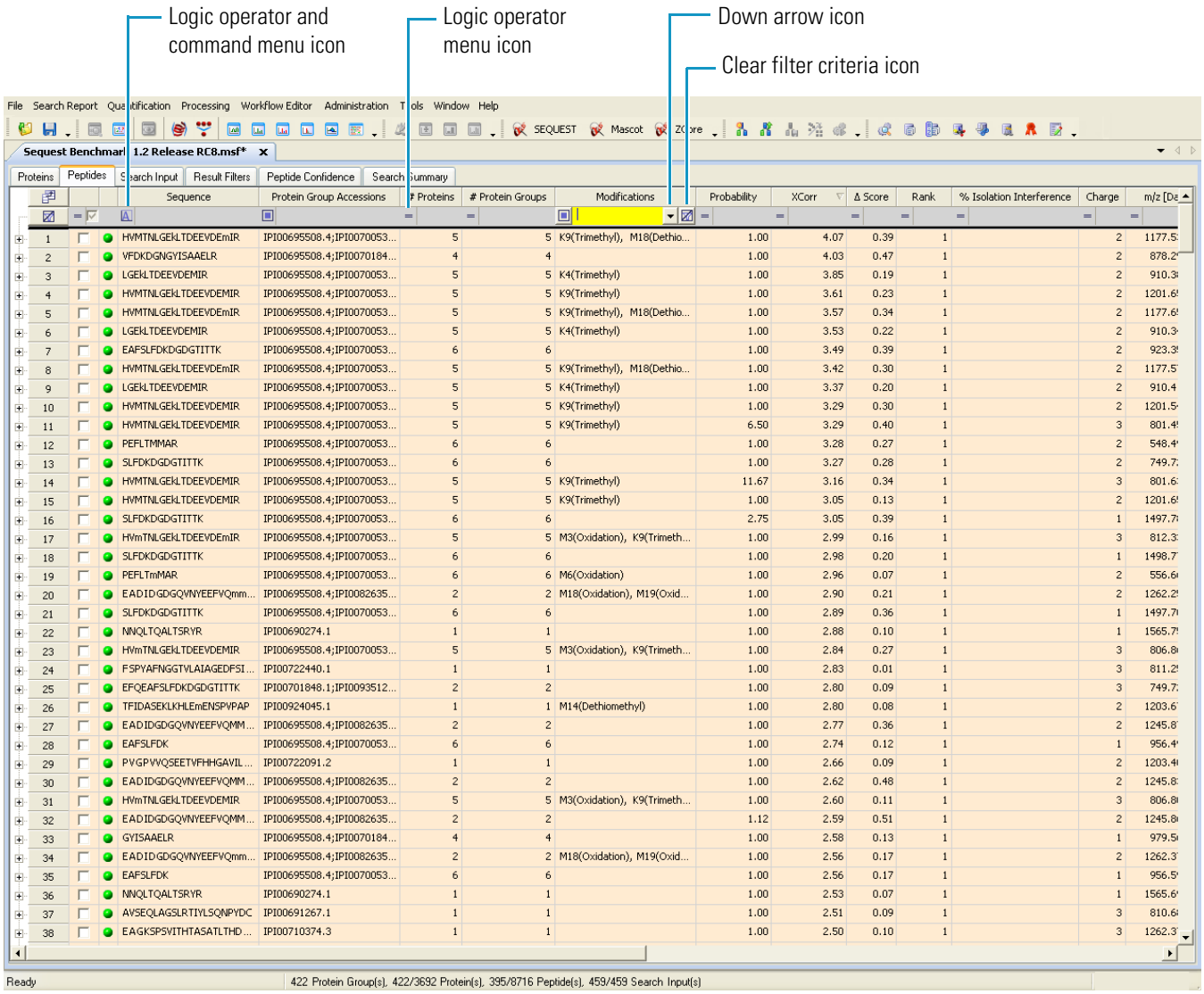


Table 65 describes the row filter icons.

Table 65. Row filter icons

Icon	Function
	This logic operator and command menu icon appears in columns that contain text. It opens the menu shown in Figure 204, which contains the following commands: • Equals • Does not equal • Less than • Less than or equal to • Greater than • Greater than or equal to • Like • Matches regular expression • Starts with • Contains • Ends with • Does not start with • Does not contain • Does not end with • Does not match • Not Like
	The logic operator menu icon appears in the columns that contain only numbers. It opens the menu shown in Figure 205, which contains the following logic operators: • Does not equal • Less than • Less than or equal to • Greater than • Greater than or equal to
	The down arrow icon appears when you click either  (command menu icon) or  (logic operator icon), or move the cursor over the icon. Clicking the down arrow opens the filter menu, which contains the following commands: • Custom: Opens the Enter Filter Criteria for *Header_name* dialog box, shown in Figure 207. • Blanks: Filters out rows that have data-filled cells in the column whose icon you clicked. • NonBlanks: Filters out rows that have empty cells in the column whose icon you clicked.
	The Clear Filter Criteria icon clears all filter conditions set with the row filters.

Figure 204. Logic operator and command menu

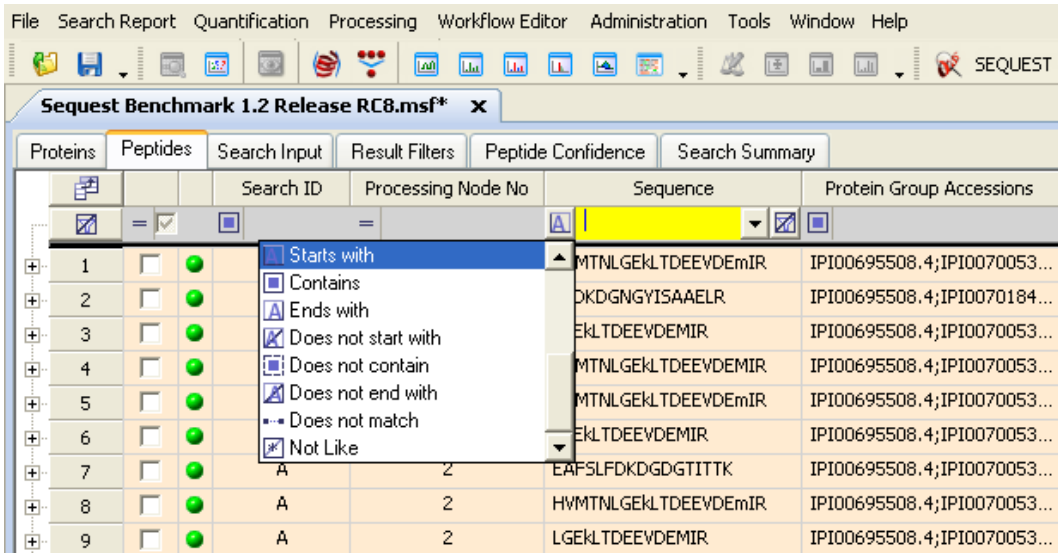


Figure 205. Logic operator menu

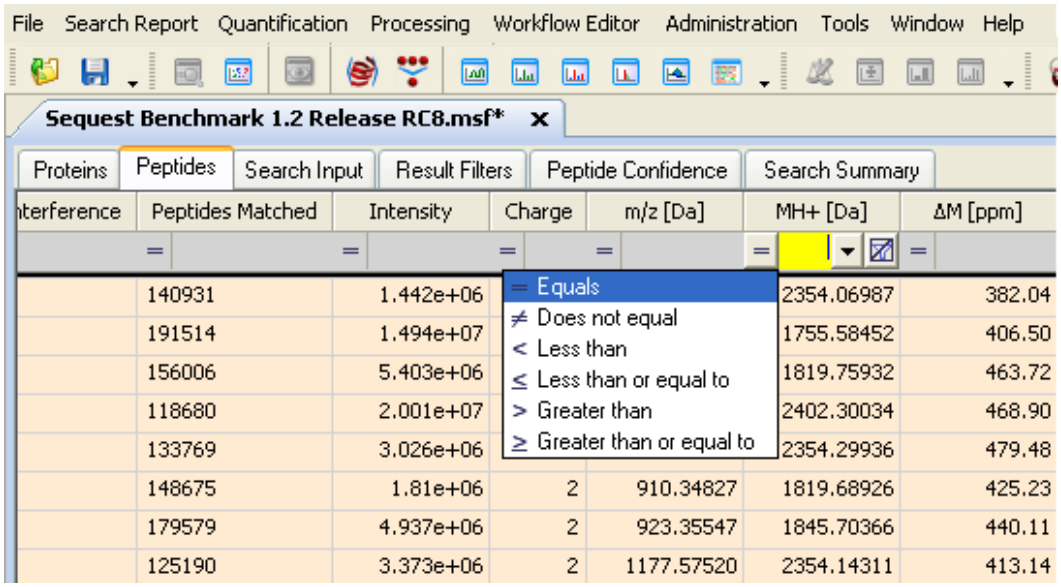
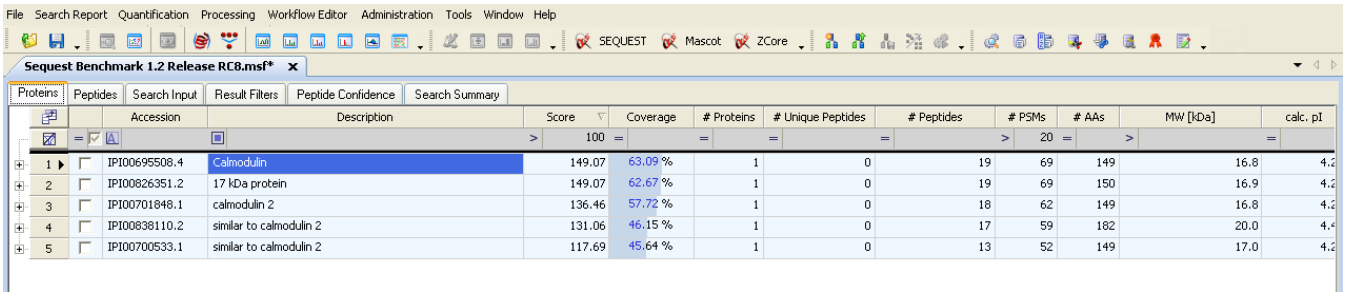


Figure 206 gives an example of simple filter criteria being entered in the row filter line. In this example, Score is set to be greater than 100, and # PSMs is set to be greater than 20.

Figure 206. Setting row filter criteria



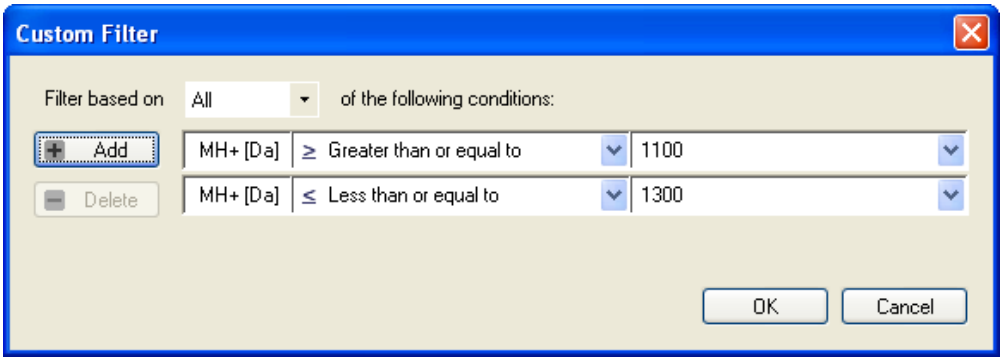
The following example shows how to use the row filter menu opened by the down arrow icon in the MSF report columns, , and the Enter Filter Criteria for *Header_name* dialog box. This example sets a precursor mass filter.

❖ To set a precursor mass filter by using the row filter menu

1. Click the down arrow icon, , and choose **Custom** from the menu.

   The Custom Filter dialog box appears, as shown in Figure 207.

   Figure 207. Custom Filter dialog box

2. From the list in the center, select the logic operator value, for example, **≥ Greater than or equal to**.
3. In the box to the right, type a value, for example, **1100**.
4. To open another row in the Custom Filter dialog box, click **Add**.
5. From the list in the center, select the logic operator value, for example, **≤ Less than or equal to**.
6. In the box to the right, type a value, for example, **1300**.

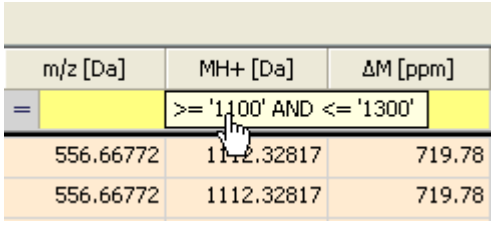
   The Custom Filter dialog box should look like the example in Figure 207.

7. Click **OK** to accept the filter settings.

   In this example, only peptides that have a precursor MH+ mass between 1100 and 1300 are displayed.

The filter conditions that you set appear when you move the cursor over the filters row, as shown in Figure 208.

Figure 208. Displaying filter conditions



❖ To clear all filter conditions set by the row filter menu

- Click the **Clear Filter Criteria** icon, , if you want to clear all filter criteria set by the commands on the row filter menu (opened by clicking the down arrow icon,).

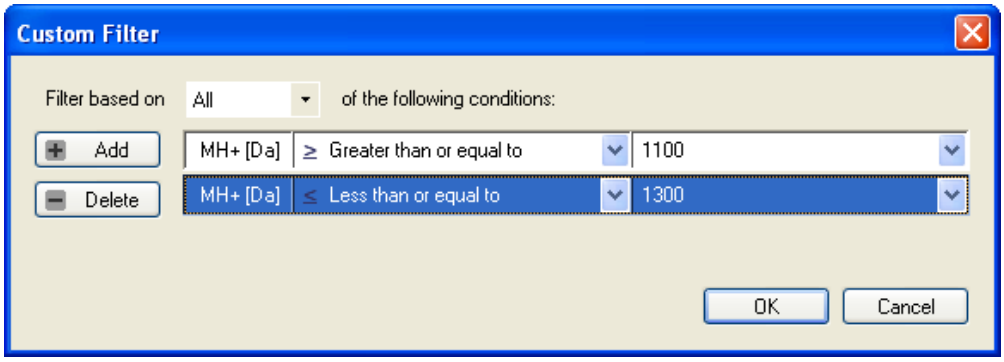
❖ To clear an individual filter set by the row filter menu

1. In the appropriate column, move your cursor over the row with the filter set by the commands on the row filter menu (opened by clicking the down arrow icon,).
2. Click the down arrow icon, , and choose **Custom**.

   The Custom Filter dialog box appears.
3. Click in the first column in the row of interest.

   The condition is activated, as shown in Figure 209.

   Figure 209. Deleting filter condition



4. Click **Delete**.
5. Click **OK**.

Custom Filter Dialog Box Parameters

Table 66 describes the parameters in the Custom Filter dialog box.

Table 66. Custom Filter dialog box parameters

Parameter	Description
And conditions	Applies AND logic to the specified conditions.
Or conditions	Applies OR logic to the specified conditions.
Operator	Specifies the operator of the filter. You can choose from the following commands: • Equals • Does not equal • Less than • Less than or equal to • Greater than • Greater than or equal to • Like • Matches regular expression • Starts with • Contains • Ends with • Does not start with • Does not contain • Does not end with • Does not match • Not Like
Operand	Specifies the operand of the filter. You can choose from the following commands or enter a value: • Blanks: Filters out rows that have data-filled cells. • DBNULL: Is only a placeholder. • Empty Text: Is only a placeholder. • []: Is only a placeholder. • Column header names
Add a condition	Opens a Operator/Operand line, so you can add a condition.
Delete Condition	Deletes the condition that you have activated with the condition button.

Displaying Filtered-Out Rows

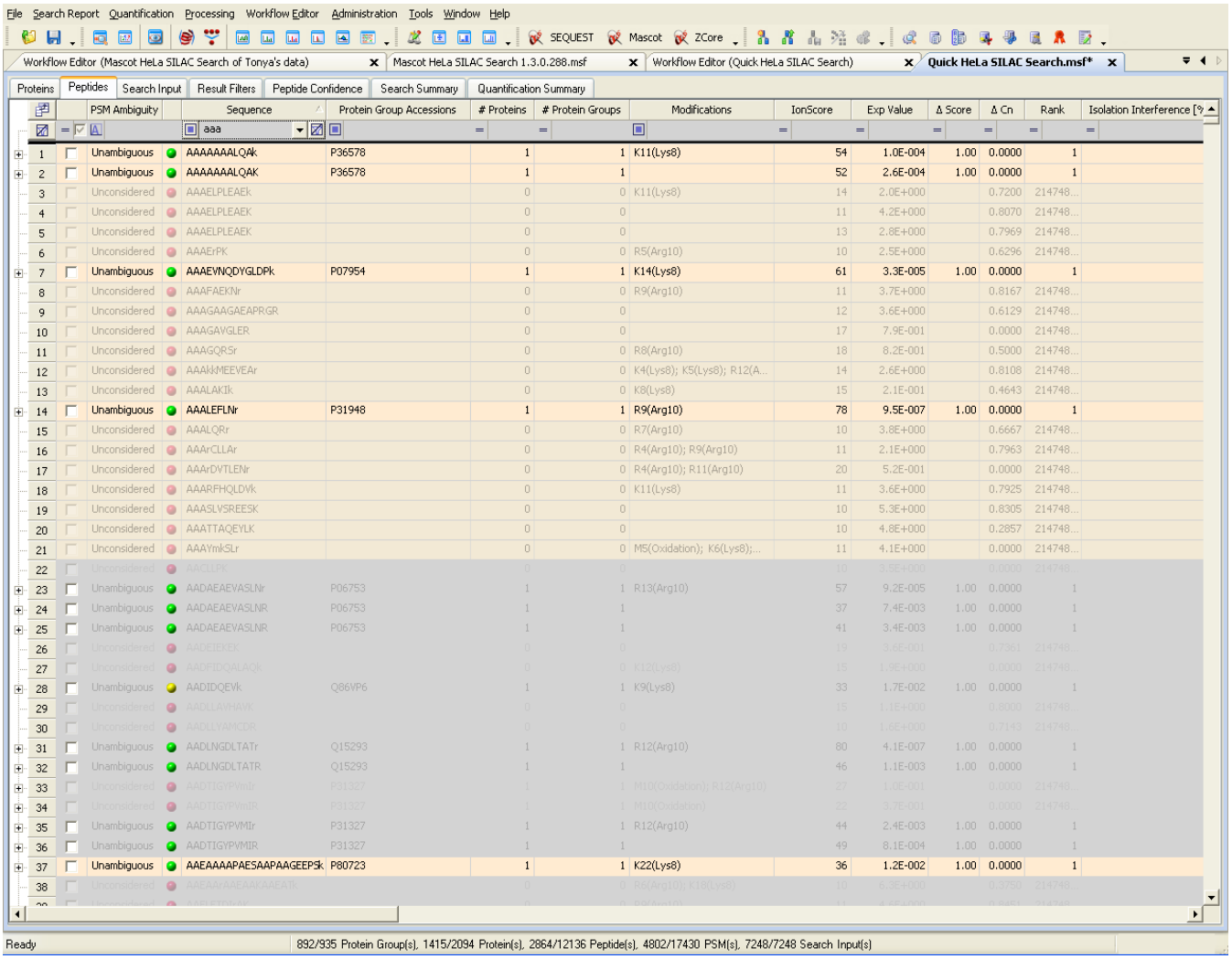
If you choose Enable Row Filters in the shortcut menu, the Proteome Discoverer application hides the filtered-out rows on the Proteins, Peptides, or Search Input page so that you can easily view your results. However, you can still display these filtered-out rows to perform a comparative analysis.

❖ To display filtered-out rows

1. Follow the procedure given in "Filtering Results with Row Filters" on page 393 to set any row filters.
2. Right-click to display the shortcut menu, as shown in Figure 210, and choose **Show Filtered Out Rows**.

   The application now displays both the filtered-out and unfiltered rows. The peptides or proteins filtered out by filters set on the Result Filters page appear in light gray rows. The peptides or proteins filtered out by row filters appear in a darker gray rows. Figure 210 shows both types of filtered-out rows.

Figure 210. Displaying filtered-out rows



3. To hide the filtered-out rows after you have displayed them, right-click the page and again choose **Show Filtered Out Rows**.

Grouping Proteins

Although MS/MS-based proteomics studies are centered around peptides, you can also explore what proteins are present in a sample and their associations through related peptides. Deducing protein identities from a set of identified peptides becomes difficult because of sequence redundancy, such as the presence of proteins that have shared peptides. These redundant proteins are automatically grouped and are not initially displayed in the search results report.

In the results report, you can turn protein grouping on or off with the Enable Protein Grouping command on the shortcut menu or with the settings in the Protein Grouping (Enabled) area on the Result Filters page. The latter method enables you to select more options in grouping. Grouping is turned on by default. For information about the grouping mechanism that the Proteome Discoverer application uses to group proteins, see "Protein Grouping Algorithm" on page 407.

The proteins within a group are ranked according to the number of peptide sequences, the number of PSMs, their protein scores, and the sequence coverage. The top-ranking protein of a group becomes the master protein of that group. By default, only the master proteins are displayed on the Proteins page.

Proteins are grouped according to the peptide sequences identified for the proteins. A protein group consists of the following:

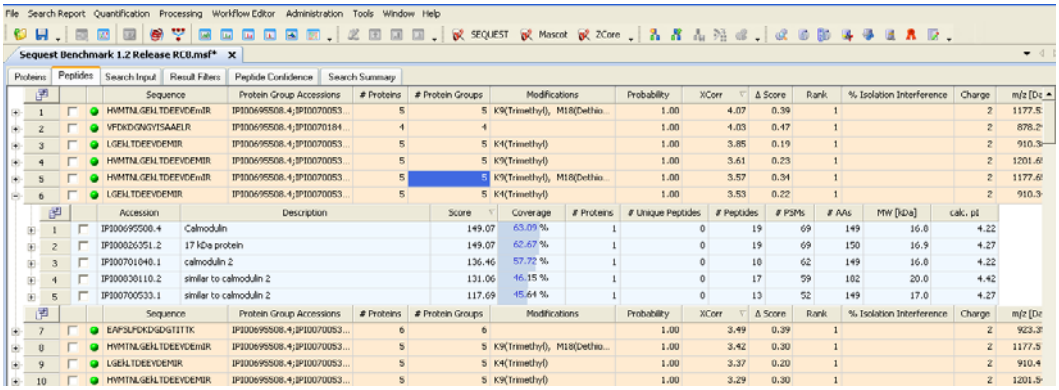
- One master protein that is identified by a set of peptides that are not included (all together) in any other protein group
- All proteins that are identified by the same set or a subset of those peptides

The # Proteins column on the Proteins and Peptides pages of the results report displays the number of identified proteins in the protein group of a master protein. It should match the number of proteins that are displayed in the Protein Group Members view when you choose Search Report > Show Protein Group Members (see Figure 213 on page 406).

Protein groups can overlap because proteins might be included in several master proteins. Each of two compared master proteins must have at least one peptide that is not contained in the other master protein. However, if you do not select the Apply Strict Maximum Parsimony Principle option in the Protein Grouping area of the Result Filters page, the peptides that distinguish these two master proteins could be contained in other master proteins. A master protein does not have to contain a unique peptide, unless you select the Apply Strict Maximum Parsimony Principle option. A unique peptide is only contained in the proteins of one protein group. In the results report, the # Unique Peptides column on the Proteins page displays the number of distinct peptide sequences for a protein group.

When you expand an identified peptide, as shown in Figure 211, the Peptides page shows only the master proteins of all protein groups that contain the peptide. To display all the proteins that belong to any of the protein groups, choose **Search Report > Show Protein Group Members View**, which opens the Protein Group Members view (see Figure 213 on page 406). To display all proteins that contain the peptide, choose **Search Report > Show Protein References**, which opens the Protein References of a Peptide view (see Figure 157 on page 287). The # Unique Peptides column on the Proteins page displays the number of peptide sequences unique to a protein group.

Figure 211. Expanding an identified peptide



Go to the following sections:

- To group the proteins in your search results and set grouping options
- To display other proteins belonging to the same protein group
- To turn off protein grouping

❖ To group the proteins in your search results

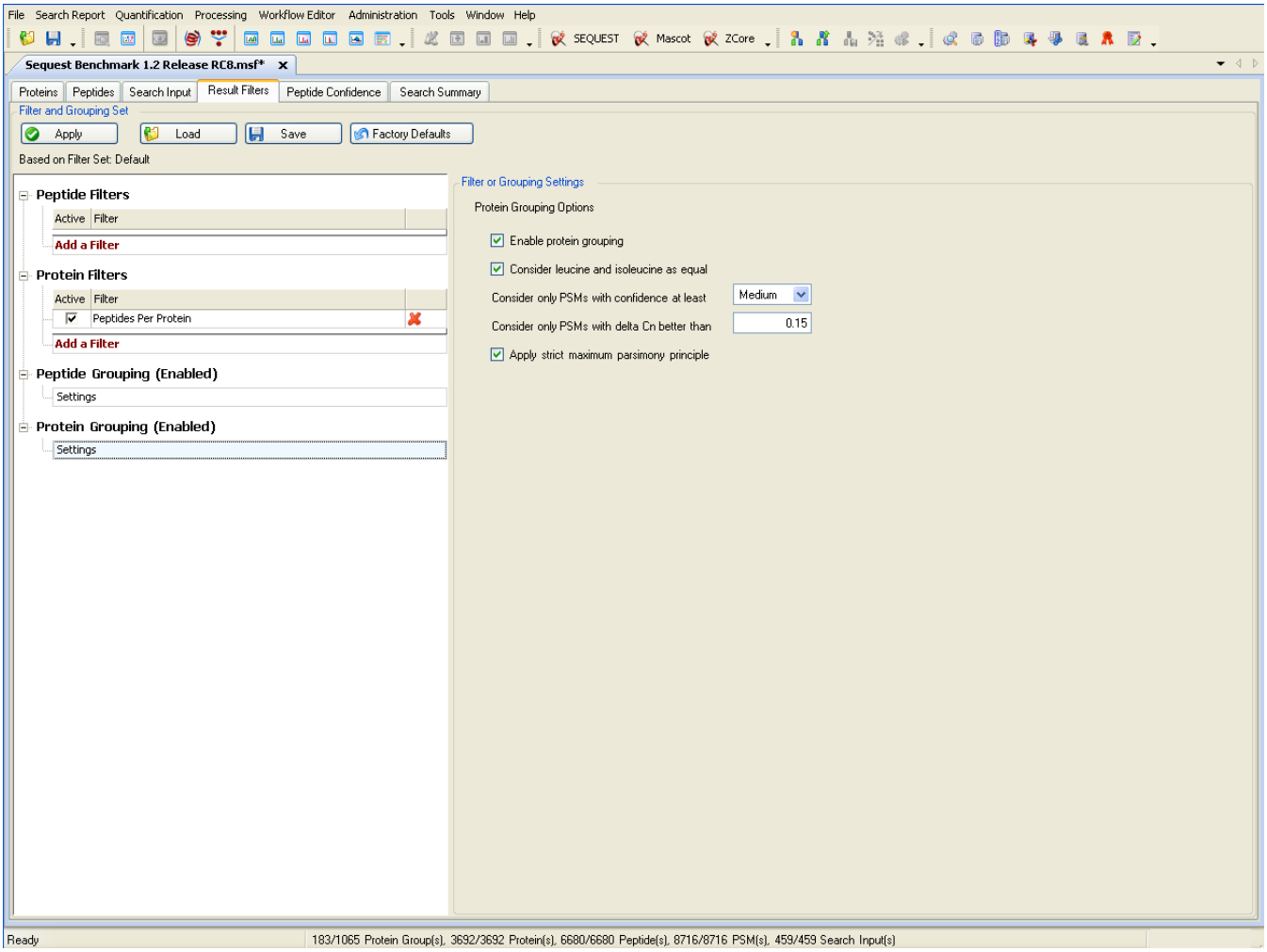
1. Open the MSF file.
2. On the Peptides or Proteins page of the MSF file, right-click a protein grid cell or row to access the shortcut menu, and choose **Enable Protein Grouping**.

❖ To group the proteins in your search results and set grouping options

1. Open the MSF file.
2. Click the **Result Filters** tab.
3. On the Results Filters page, click **Settings** beneath Protein Grouping (Enabled).

   Protein grouping options appear in the Filter or Grouping Settings area, as shown in Figure 212.

Figure 212. Protein grouping options



4. If you want to group homologous proteins, select the **Enable Protein Grouping** check box, if it is not already selected by default.

5. If you want leucine and isoleucine to be treated as equal, select the **Consider Leucine and Isoleucine as Equal** check box, if it is not already selected by default.

   The Proteome Discoverer application groups similar proteins together. Proteins are similar if their peptides are similar. If you select the Consider Leucine and Isoleucine as Equal check box, the application treats leucine and isoleucine as equal amino acids.

6. To specify the type of PSMs that the Proteome Discoverer application considers for inclusion in protein grouping, set the **Consider Only PSMs with Confidence at Least** parameter to the desired setting:
   - Low: Considers all (low-, medium-, and high-confidence) PSMs for inclusion in protein grouping.
   - (Default) Medium: Considers medium- and high-confidence PSMs for inclusion in protein grouping.
   - High: Considers high-confidence PSMs for inclusion in protein grouping.
7. If you want the Proteome Discoverer application to consider only PSMs with values lower than or equal to a specified value for inclusion in the protein grouping process, specify a value in the **Consider Only PSMs with Delta Cn Better Than** box.

   The default Delta Cn value is 0.15. To have the Proteome Discoverer application consider all PSMs, set the value to 1.0.
8. If you want to remove all protein groups that are not necessary to explain the found peptides, select the **Apply Strict Maximum Parsimony Principle** check box.

   The Apply Strict Maximum Parsimony Principle option ensures that only one PSM per spectrum is used for protein grouping. If the Delta Cn range of the spectrum includes more than one PSM, the Proteome Discoverer application selects the "best" PSM and rejects the others for grouping and quantification.
9. Click 

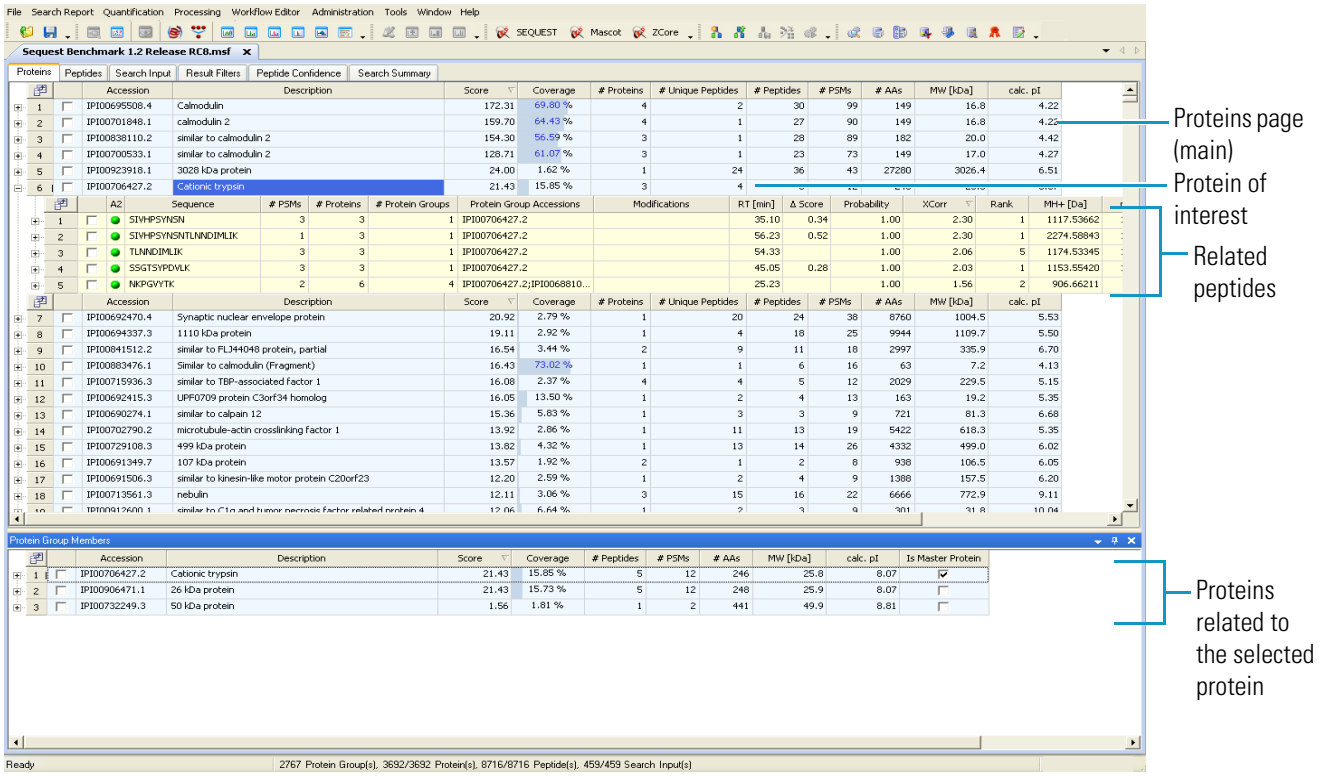
.

❖ To display other proteins belonging to the same protein group

1. Open the MSF file.
2. On the Proteins page, click anywhere in a protein row.
3. Choose **Search Report > Show Protein Group Members View**, or click the **Show Protein Group Members view** icon,  .

   The Protein Group Members view appears below the Proteins page, as shown in Figure 213.

Figure 213. Proteins in the same group



The Is Master Protein column in the Protein Group Members view indicates whether the protein is the master protein of a protein group. For some peptides, a list of proteins might contain this peptide sequence, but none of them is a master protein. This situation can occur if the peptide contains isoleucine at a position where the master protein has leucine, or vice versa, and the proteins are grouped with the Consider Leucine and Isoleucine as Equal option on the Result Filters page.

❖ To turn off protein grouping

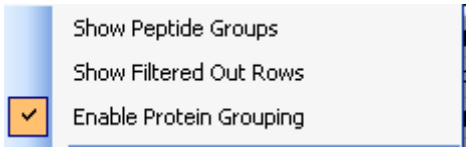
1. On the Result Filters page, click on **Settings** below Protein Grouping (Enabled), and clear the **Enable Protein Grouping** check box.

   –or–

   On the Proteins or Peptides page, right-click a protein grid cell or row to access the shortcut menu, and clear the check mark for **Enable Protein Grouping**, shown in Figure 214.

   The proteins are no longer grouped.

   Figure 214. Enable Protein Grouping command on the Proteins page shortcut menu



2. To regroup proteins, reselect the **Enable Protein Grouping** check box on the Result Filters page.

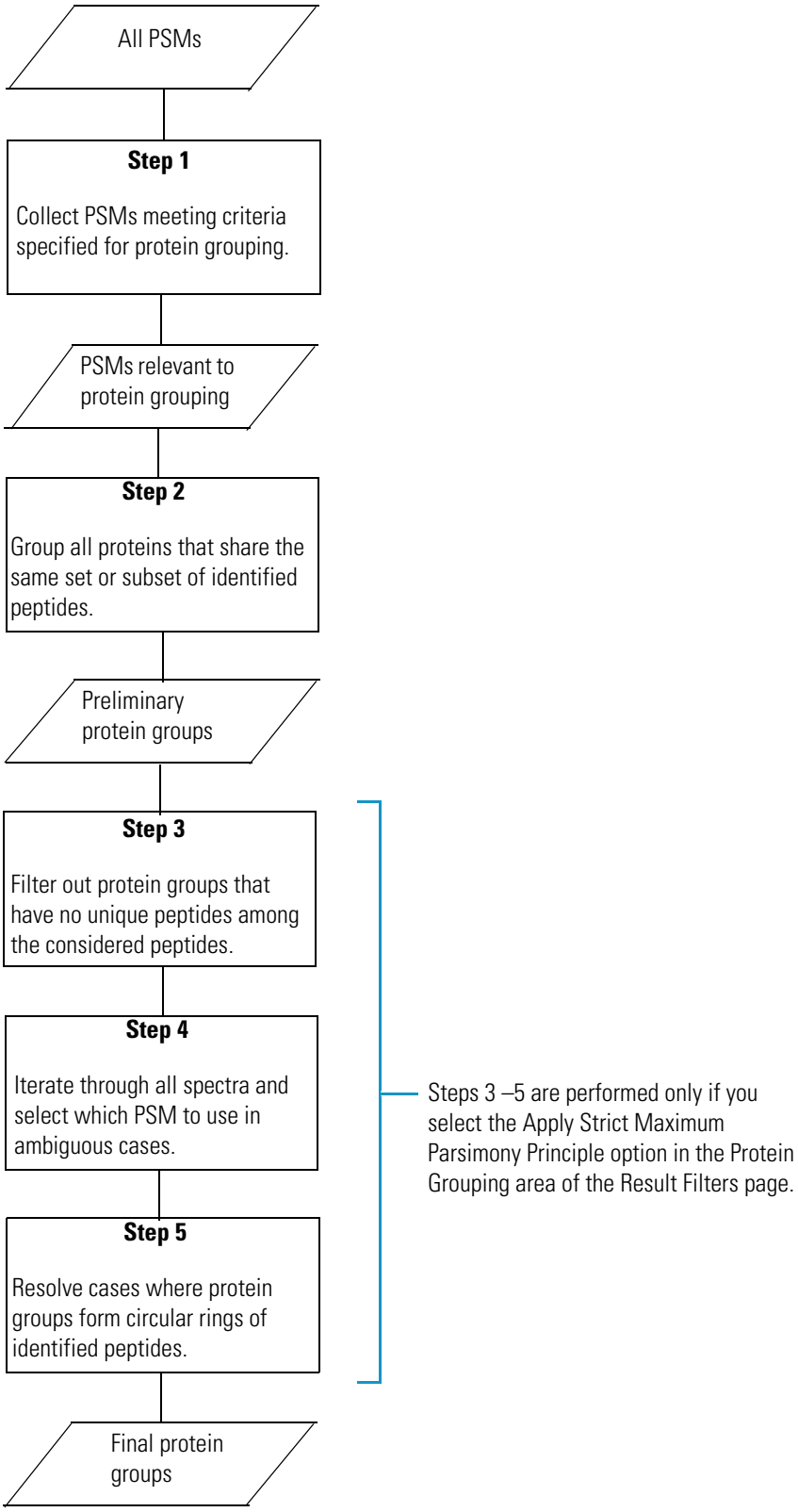
   –or–

   Right-click a protein grid cell or row in the Proteins or Peptides page and choose **Enable Protein Grouping** from the shortcut menu.

Protein Grouping Algorithm

The Proteome Discoverer application uses a protein grouping inference process to group proteins. Figure 215 shows the steps involved in this process.

Figure 215. Protein grouping inference process in the Proteome Discoverer application



1. In the first step, the application collects all peptide spectrum matches (PSMs) that meet the selection criteria that you specified through the settings of the parameters in the Protein Grouping (Enabled) area on the Result Filters page (see Figure 212 on page 404). Table 59 explains these parameters. You can use these settings to specify which PSMs to consider for the inference of the protein groups. For example, if you set the Consider Only PSMs with Confidence at Least parameter to Medium, the Proteome Discoverer application considers only PSMs with a medium- or high-identification confidence when it creates the protein groups and ignores PSMs with a low-identification confidence. You can further use the Consider Only PSMs with Delta Cn Better Than parameter to filter out PSMs over a normalized score and consider the remaining PSMs for inclusion in the protein group inference process if their confidence levels fit.

   Note Setting the Consider Only PSMs with Confidence at Least parameter to Low and the Consider Only PSMs with Delta Cn Better Than parameter to 1 and leaving the Apply Strict Maximum Parsimony Principle option unselected creates the same protein groups as the previous release of the Proteome Discoverer application.

   This first step prevents protein groups from including low-scoring, low-confidence PSMs. Even if the Proteome Discoverer application loads all PSMs initially identified by the search engines without applying further result filters, it considers only those PSMs meeting the specified criteria when inferring protein groups. If the set result filters filter out PSMs, the application does not consider them for the protein grouping process, even if they would otherwise fit the set grouping criteria.

2. In the second step, the application creates preliminary protein groups from the PSMs collected in the first step. It combines all proteins into one protein group that contains the same subset of peptides.

   The Proteome Discoverer application takes the next steps in the protein grouping process if you select the Apply Strict Maximum Parsimony Principle parameter in the Result Filters page.

3. In the third step, the application removes all protein groups that have no unique peptides among the peptides that it considers for the protein grouping process. If a protein group does not contain at least one unique peptide, all of its peptides are also included by other protein groups, so there is no supporting evidence for the existence of this protein group. At this point, the application explicitly retains all protein groups that form circular rings of overlapping shared peptides. For example, suppose a circular ring is composed of the protein groups:

   - ABCD (identified by peptides a, b, c, and d)
   - CDEF (identified by peptides c, d, e, and f)
   - EFAB (identified by peptides e, f, a, and b)

   To explain all identified peptides, only two of the three protein groups are needed, but at this point it is not clear which to take and which to reject. The application postpones the resolution of this issue until step 5.

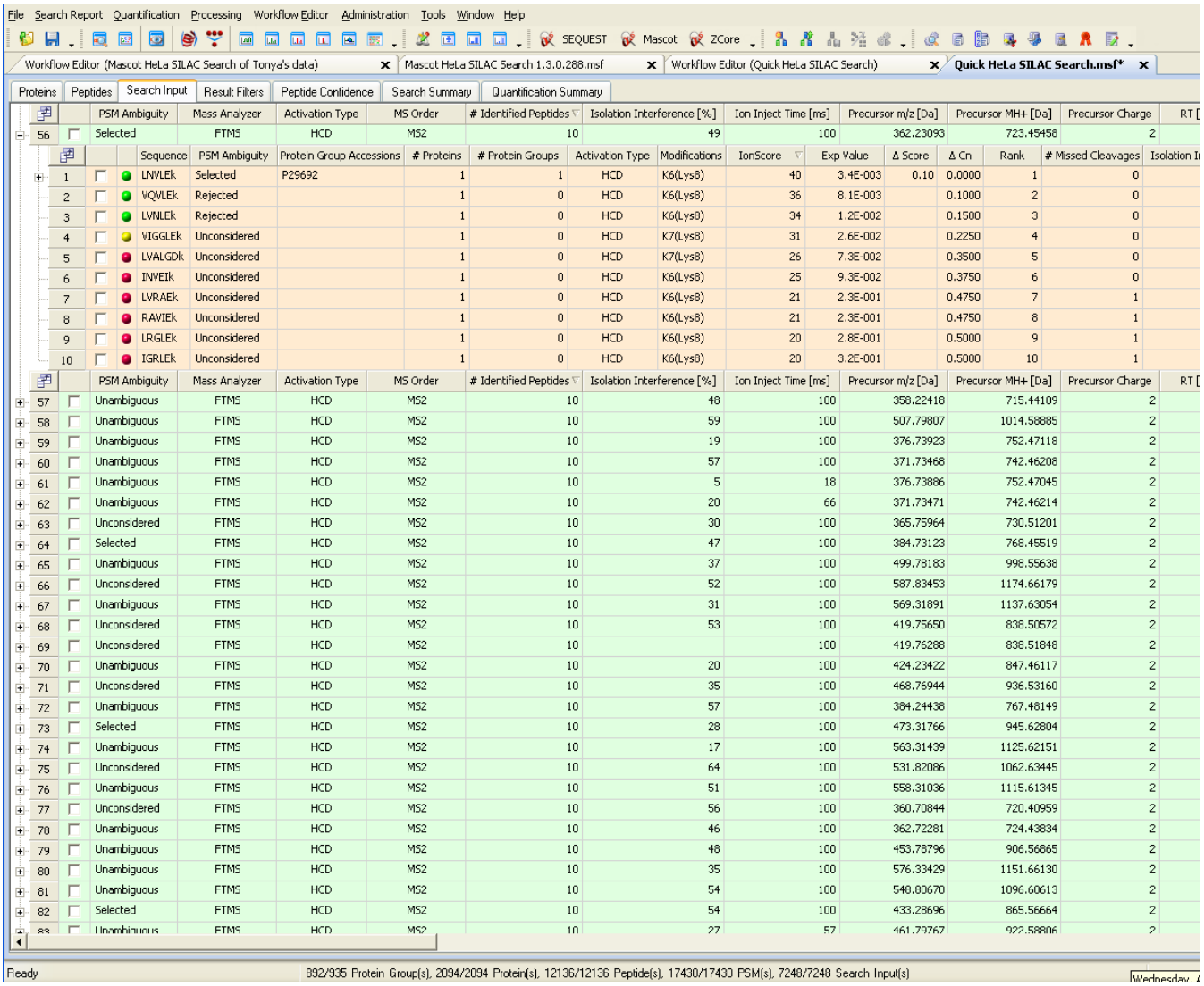
4. In the fourth step, the application first collects all spectra with more than one peptide match to consider for the protein grouping process. It then resolves these ambiguous cases and selects one of the PSMs to use for the protein grouping process while rejecting the remaining peptide matches of a spectrum. In cases where more than one PSM is considered for a spectrum, it resolves this ambiguity by selecting the PSM that is connected to the "best" protein group and rejecting the other PSMs. The "best" protein group is the group with the highest number of unambiguous and unique peptides and the highest protein score.
5. In the fifth step, the application resolves the cases where protein groups form circular rings of overlapping identified peptides. This step is the last step of the protein group inference process, resulting in the final list of protein groups that are reported in the Proteins page of the MSF file.

The PSM Ambiguity column on the Peptides and Search Input pages can help you understand the process of selecting PSMs for the protein group. This column is available for every PSM, every search input entry (representing the searched spectra), and every peptide group. For the search input entries and the peptide groups, this column displays the best PSM ambiguity from all connected PSMs. See Table 45 for a description of the categories of ambiguity in this column.

Note If you want to investigate the protein grouping mechanism in detail, set the Group Peptides By option in the Peptide Grouping (Enabled) area of the Result Filters page to Sequence and not to Mass and Sequence. This way, the peptide groups created are similar to the protein groups created, which are always based on peptide sequences.

Consider the example shown in Figure 216, where 10 different PSMs are identified for search input 56. The three PSMs ranked 1 through 3 all meet the specified protein grouping criteria. They are of high confidence, and their Delta Cn values are below the threshold of 0.15, so the protein group inference algorithm considers all three PSMs for grouping. It does not consider the remaining PSMs of the spectrum, which are ranked 4 through 10 and are of low confidence, when creating protein groups.

Figure 216. PSMs shown for search input



Proteins Containing Peptides with Sequences Not Belonging to Master Protein

Because the Proteome Discoverer application considers for inclusion in the protein grouping process only PSMs that meet the criteria set in the Protein Grouping (Enabled) area of the Results Filters page, a protein group might contain proteins that have identified peptides whose sequences are not all contained in the master protein of the protein group. For example, if you specify that the protein grouping inference process consider only PSMs that have at least medium confidence, a protein group might include a protein with a low-confidence peptide that does not belong to a master protein.

Protein Groups in the Status Bar

The status bar shows the actual number of protein groups versus the total number of protein groups (see Figure 9 on page 27). The difference is the number of protein groups that the application removed to comply with the selection of the Apply Strict Maximum Parsimony Principle option on the Results Filters page. Enabling the display of filtered-out protein groups allows investigating the protein groups that were removed during this process.

Proteins Grouped by the Grouping Algorithm in Previous Releases

The Proteome Discoverer application removes some protein groups that the protein grouping mechanism created in previous versions of the application. The previous algorithm might have created these groups from only low-confidence peptides, or the application removed them to comply with the selection of the Apply Strict Maximum Parsimony Principle option on the Results Filters page. Some peptides might therefore not belong to any protein group. To investigate these cases, right-click on the Proteins page and select **Show Filtered Out Rows** to display the filtered-out peptides in the results file. You can also use the Protein References of a Peptide view opened by the Search Report > Show Protein References command to help you.

Number of Unique Peptides Column on Proteins Page

The value in the # Unique Peptides column on the Proteins page that is listed for each protein group is the number of peptides that are only contained in this protein group. The Proteome Discoverer application counts only peptides that display a status of Selected or Unambiguous in the PSM Ambiguity column, because it makes no sense to assess the uniqueness of peptides that were not used to form protein groups.

PSMs Identified by Multiple Workflow Nodes

In search results where the application identifies PSMs by multiple search nodes within a single workflow, the protein grouping algorithm selects one of the PSMs identified for the same spectrum for building the protein groups.

In search results where PSMs are identified by multiple search nodes from multiple workflows (multiconsensus report), the application treats PSMs and spectra from the different workflows as being separate, even if it searched the same raw data files and therefore the same spectra. In this case, it is difficult to determine whether the application searched the exact same spectra, because they might have changed in the different workflows.

Grouping Peptides

In the results report, you can turn peptide grouping on or off with the Show Peptide Groups command on the shortcut menu or with the settings in the Peptide Grouping (Enabled) area on the Result Filters page. The latter method enables you to select more options in grouping. Grouping is turned on by default.

In the Peptide Grouping area of the Result Filters page, you can specify whether you want to group peptides only by sequence or by mass and sequence. The Mass and Sequence setting of the Group Peptides By option separates the differently modified forms of a peptide into different peptide groups. This setting is the default.

The number of peptides displayed in the status bar is always the number of distinct sequences. The number of peptide groups, on the other hand, depends on the peptide grouping settings. If you group peptides by sequence only, the two numbers are the same. If you group peptides by sequence and mass, the number of peptide groups is normally larger than the number of peptides displayed in the status bar, unless the peptides have no modifications.

❖ To group the peptides in your search results

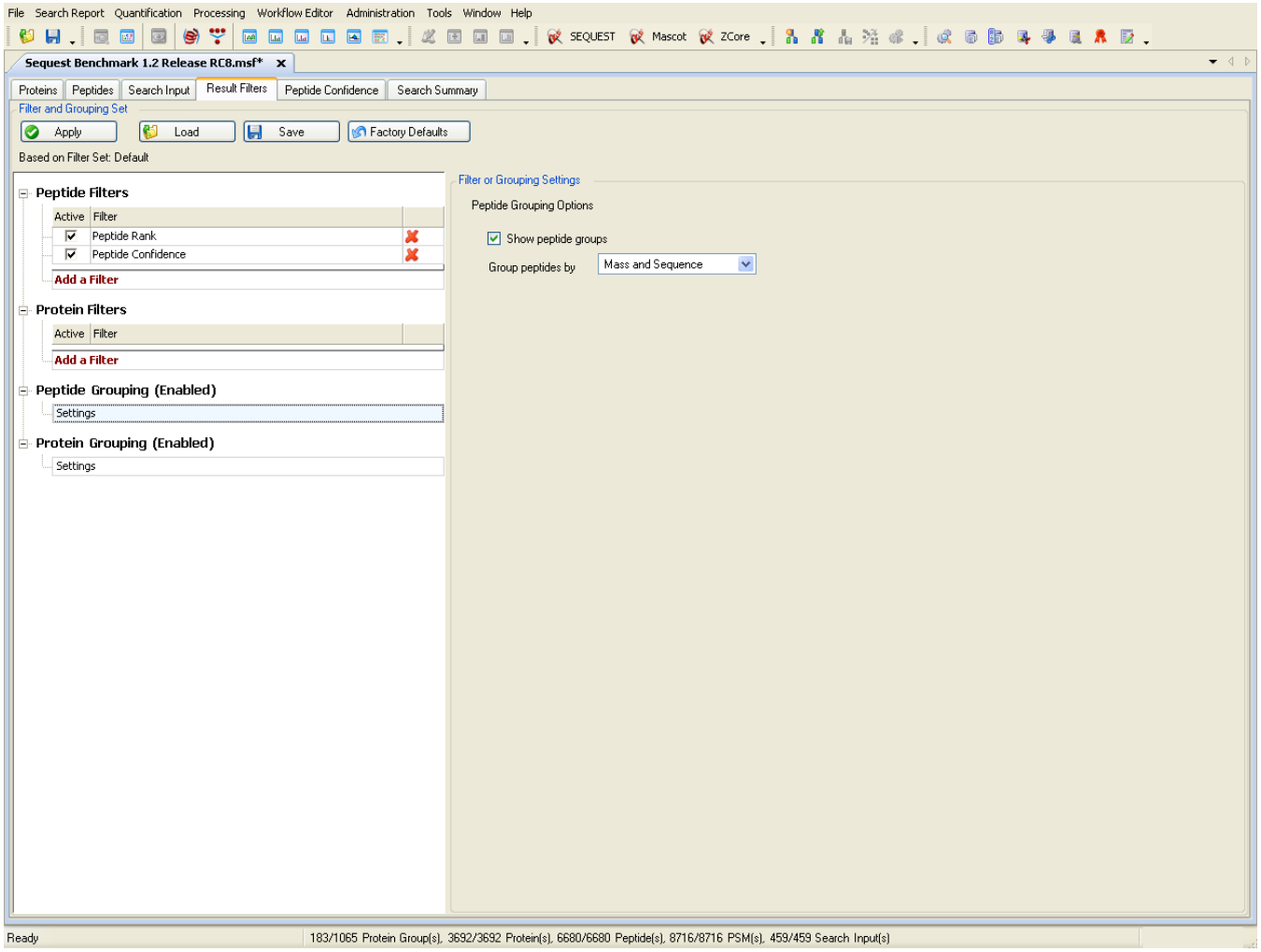
1. Open the MSF file.
2. On the Peptides or Proteins page of the MSF file, right-click a peptide grid cell or row to access the shortcut menu, and choose **Show Peptide Groups**.

❖ To group the peptides in your search results and set grouping options

1. Open the MSF file.
2. Click the **Results Filters** tab.
3. On the Results Filters page, click **Settings** beneath Peptide Grouping (Enabled).

   Peptide grouping options appear in the Filter or Grouping Settings area, as shown in Figure 217.

Figure 217. Peptide grouping options



4. If you want peptides to be grouped on the Peptides page of the results report, select the **Show Peptide Groups** check box.
5. Select the method of grouping peptides from the Group Peptides By menu:
   - Sequence: Groups peptides by sequence.
   - Mass and Sequence: Groups peptides by mass and sequence.
6. Click Apply.

Calculating False Discovery Rates

The false discovery rate (FDR), or the false positive rate, is a statistical value that estimates the number of false positive identifications among all identifications found by a peptide identification search. It is a measure of the certainty of the identification. You can use the Proteome Discoverer application's decoy database search feature to determine false discovery rates.

You can use FDRs to validate MS/MS searches of large data sets, but they are not effective on searches of a small number of spectra or searches against a small number of protein sequences, because the number of matches will likely be too small to give a statistically meaningful estimate.

A decoy database gives a probability value to identifiers and the percentage of false discoveries that you can expect. A one percent false discovery rate is a typical target for searches.

A good decoy database should contain entries that look like real proteins but do not contain genuine peptide sequences. The simplest approach to achieving such a decoy database is to reverse all protein sequences, which is the scheme that the Proteome Discoverer application currently uses. It is a suitable approach for enzymatic MS/MS searches.

CAUTION Reversing the database is not suitable for peptide mass fingerprinting or no-enzyme MS/MS searches, especially for dynamic modifications. You might see mass shifts at each end of a peptide sequence that transform a genuine y series match into a false b series match or vice versa.

You can perform the decoy database search in two ways:

- Perform two separate searches, one against the non-decoy database and one against the decoy database. Then count the number of matches from both searches to determine the false discovery rates. This approach is the more conservative approach.
- Create a concatenated database from the non-decoy and the decoy database and then perform the search against this concatenated database.

The difference between the two approaches becomes clear in the case where you find two significant matches for a given spectrum. The first match is from the non-decoy database, and the second one is from the decoy database. Because the Proteome Discoverer application considers only the top matches when calculating the false discovery rates, finding two significant matches for a given spectrum is not considered a false positive in the concatenated database approach, but it counts in the separate databases approach. The latter case is considered the more conservative one and is the approach that the application currently uses.

To calculate the false discovery rate, the application counts the matches that pass a given set of filter thresholds from the decoy database and from the non-decoy database. It counts only the top match per spectrum, assuming that for any given spectrum only one peptide can be the correct match.

Target False Discovery Rates (FDRs)

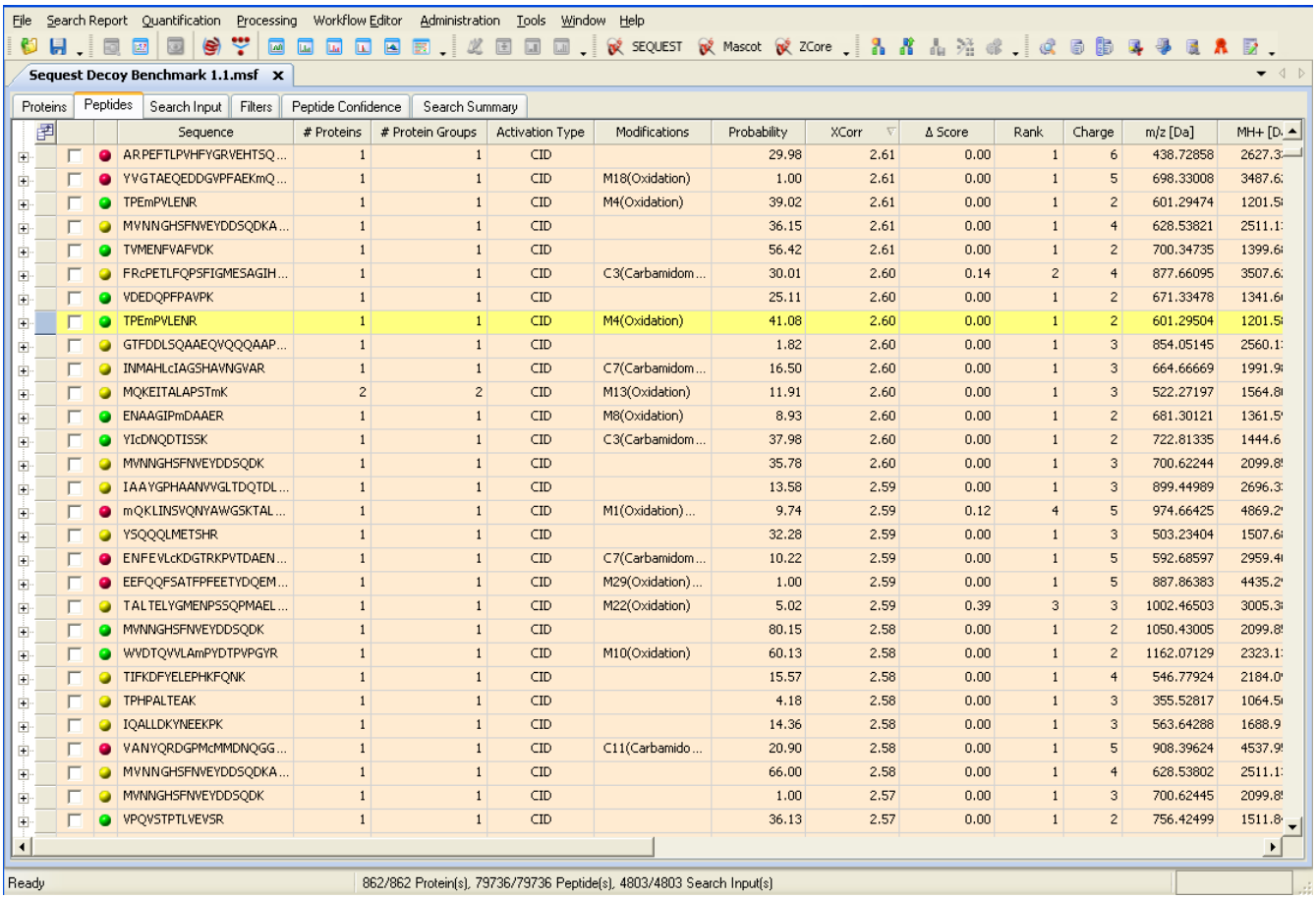
If you set a false discovery rate target value for a decoy database search, the application determines and applies filter thresholds to identified matches so that the resulting false discovery rate is not higher than the set target value. The confidence indicators applied to each peptide match are distributed according to these calculated filter thresholds (see Figure 223 on page 424).

You must specify two target values for a decoy database search: a strict target FDR and a more relaxed FDR. Figure 220 on page 420 shows the decoy search setting with target false discovery rates of one percent and five percent, respectively. After completing the search, the system automatically determines two sets of filter settings so that the resulting separate FDRs do not exceed their corresponding target value.

Peptide Confidence Indicators

The filter settings that determine false discovery rates are used to distribute the confidence indicators for the peptide matches (these are the green, yellow, and red circles attached to each peptide match). Whenever you perform a decoy database search during the database search and apply filter settings to achieve the specified target FDRs, the same filters are used to distribute the confidence indicators. Peptide matches that pass the filter associated with the strict FDR are assigned a green confidence indicator, peptide matches that pass the filter associated with the relaxed FDR are assigned a yellow confidence indicator, and all other peptide matches receive a red indicator of low confidence. Figure 218 gives an example of these confidence indicators.

Figure 218. Decoy search results



Note You can change the default confidence levels to alternative values on the Peptide Confidence page.

Setting Up False Discovery Rates in Search Wizards and the Workflow Editor

You can set up false discovery rates in both the search wizards and the Workflow Editor.

- Setting Up False Discovery Rates in the Search Wizards
- Setting Up False Discovery Rates in the Workflow Editor by Using the Peptide Validator Node
- Setting Up False Discovery Rates in the Workflow Editor by Using the Percolator Node

Setting Up False Discovery Rates in the Search Wizards

You can set the strict and relaxed false discovery rates for every available search wizard.

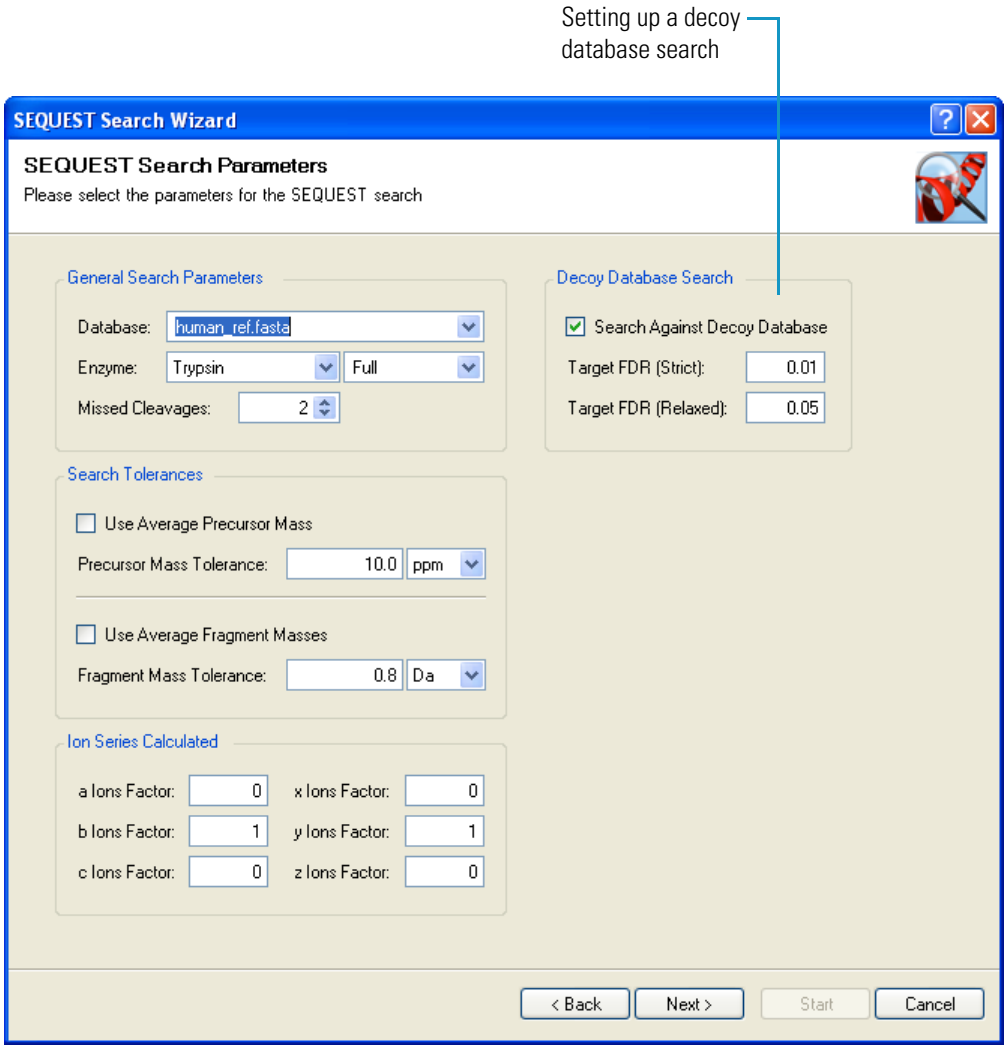
❖ To set up false discovery rates in a search wizard

1. Start your search by using the search wizards.

   For information about using the search wizards, see "Starting a New Search by Using the Search Wizards" on page 63.

2. On the <*Wizard_name*> Search Parameters page, select the **Search Against Decoy Database** option, as shown in Figure 219.

Figure 219. Setting up a decoy database search in a search wizard



3. In the Target FDR (Strict) box, set the target FDR for high-confidence peptide hits.
4. In the Target FDR (Relaxed) box, set the target FDR for medium-confidence peptide hits.
5. Click **Next**.

Setting Up False Discovery Rates in the Workflow Editor by Using the Peptide Validator Node

You can set up false discovery rates through the Peptide Validator node in the workflow.

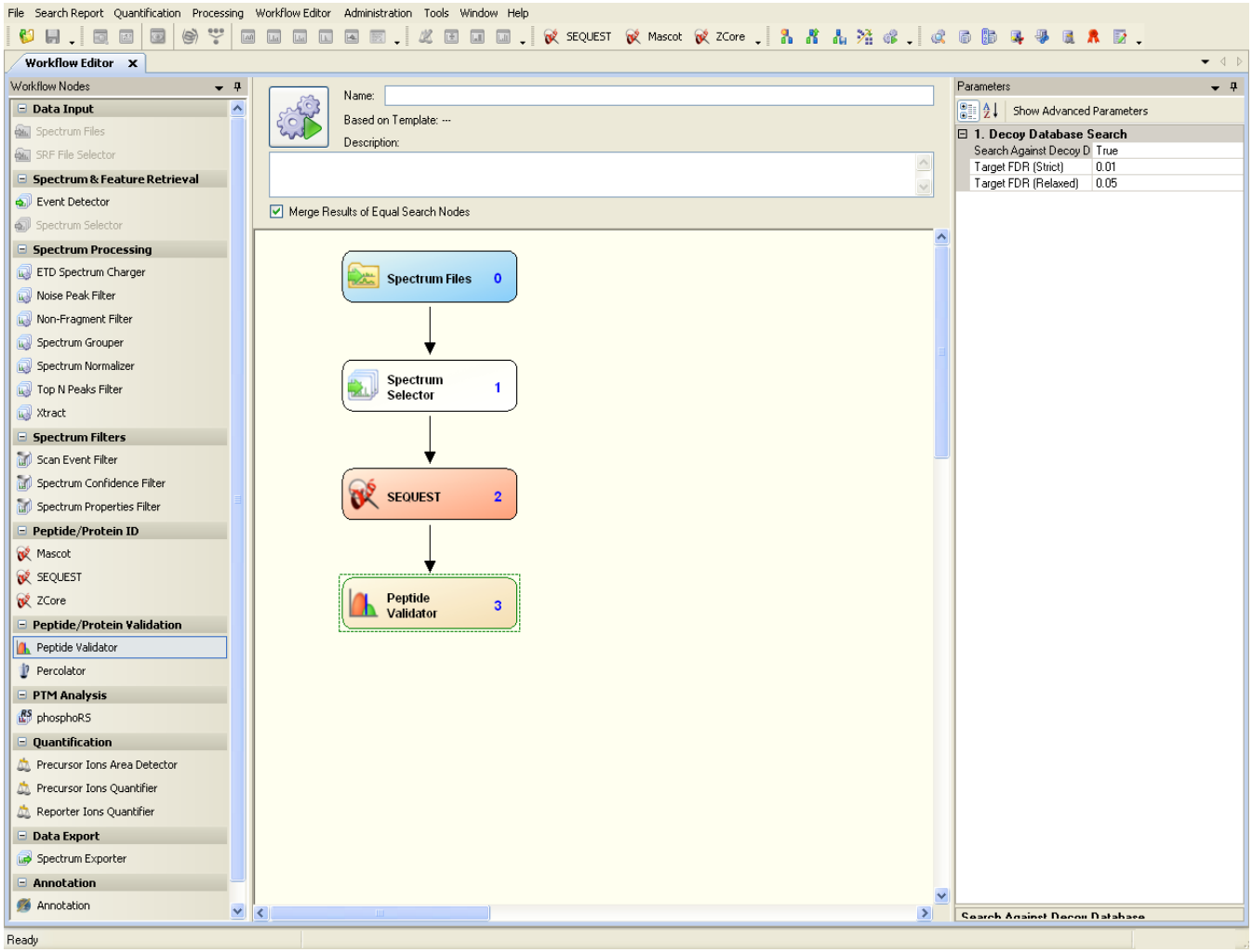
❖ To set up false discovery rates in the Workflow Editor

1. Create a search workflow that includes at least one of the search engine nodes (SEQUEST, Mascot, or ZCore) and the Peptide Validator node.

   For information about creating a workflow, see "Creating a Search Workflow" on page 91.

2. Click on the **Peptide Validator** node.
3. In the Decoy Database Search area in the Parameters pane, set Search Against Decoy Database to **True**, if it is not already set to True, as shown in Figure 220.

Figure 220. Setting up a decoy database search in the Workflow Editor



4. In the Target FDR (Strict) box, set the target FDR for high-confidence peptide hits.
5. In the Target FDR (Relaxed) box, set the target FDR for Peptides medium-confidence peptide hits.
6. Choose **Workflow Editor > Start Workflow**, or click the Start Workflow icon, .

Setting Up False Discovery Rates in the Workflow Editor by Using the Percolator Node

You can set up false discovery rates through the Percolator node in the workflow. For detailed information about the Percolator node and its processing, see "Percolator Node" on page 681.

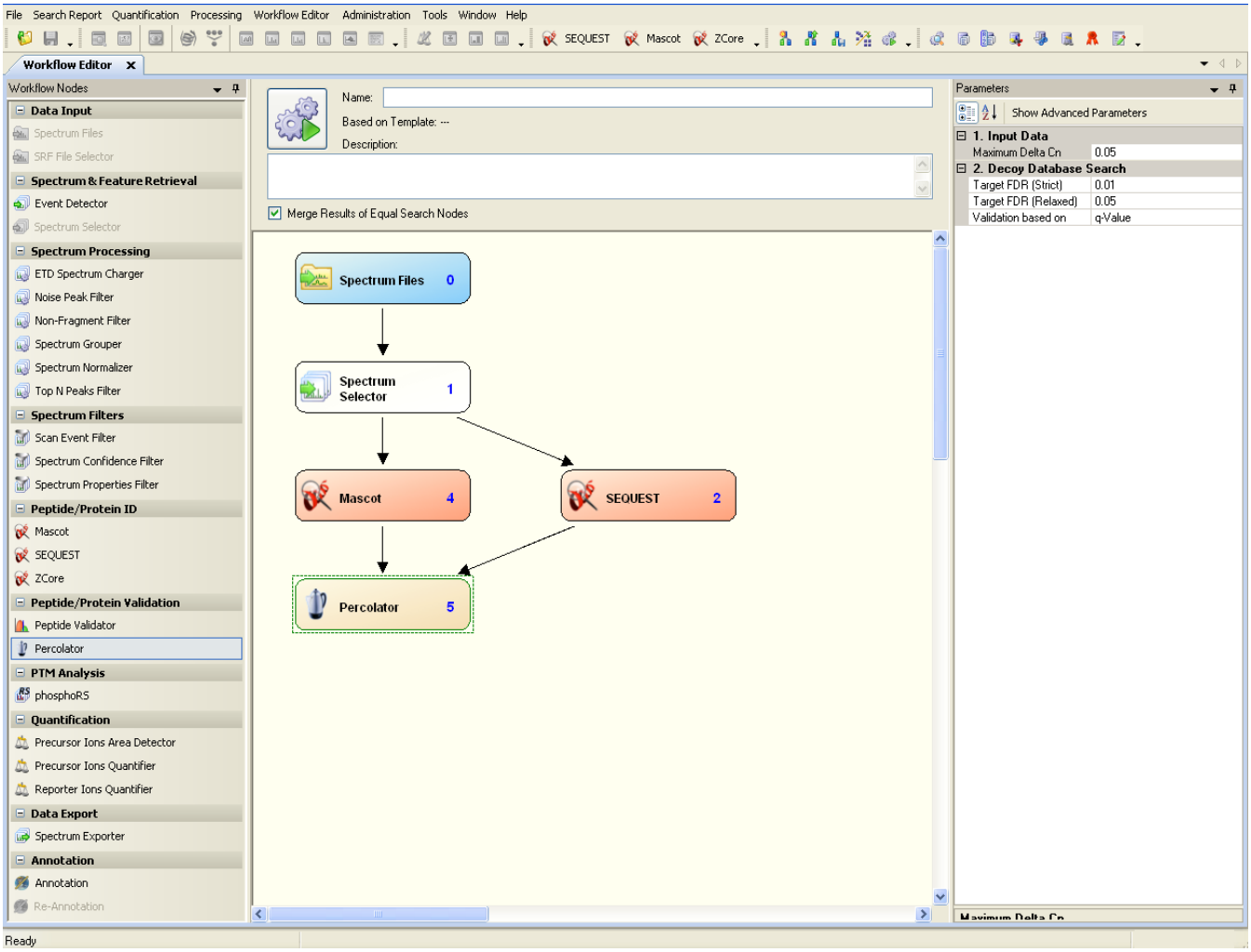
❖ To set up false discovery rates by using the Percolator node

1. Create a search workflow that includes at least one of the search engine nodes (SEQUEST, Mascot, or ZCore) and the Percolator node.
2. For information about creating a workflow, see "Creating a Search Workflow" on page 91.
3. Connect all search nodes whose results you want to submit for validation to the Percolator node.

   Figure 221 gives an example of such a workflow.

Note To work properly, Percolator needs a sufficient number of PSMs from the target and the decoy search. If the search identified fewer than 200 target or decoy PSMs, or if fewer than 20 percent decoy PSMs are available compared to the number of target matches, Percolator rejects them for processing and displays an appropriate message in the Proteome Discoverer job queue or in the Search Summary of an open report.

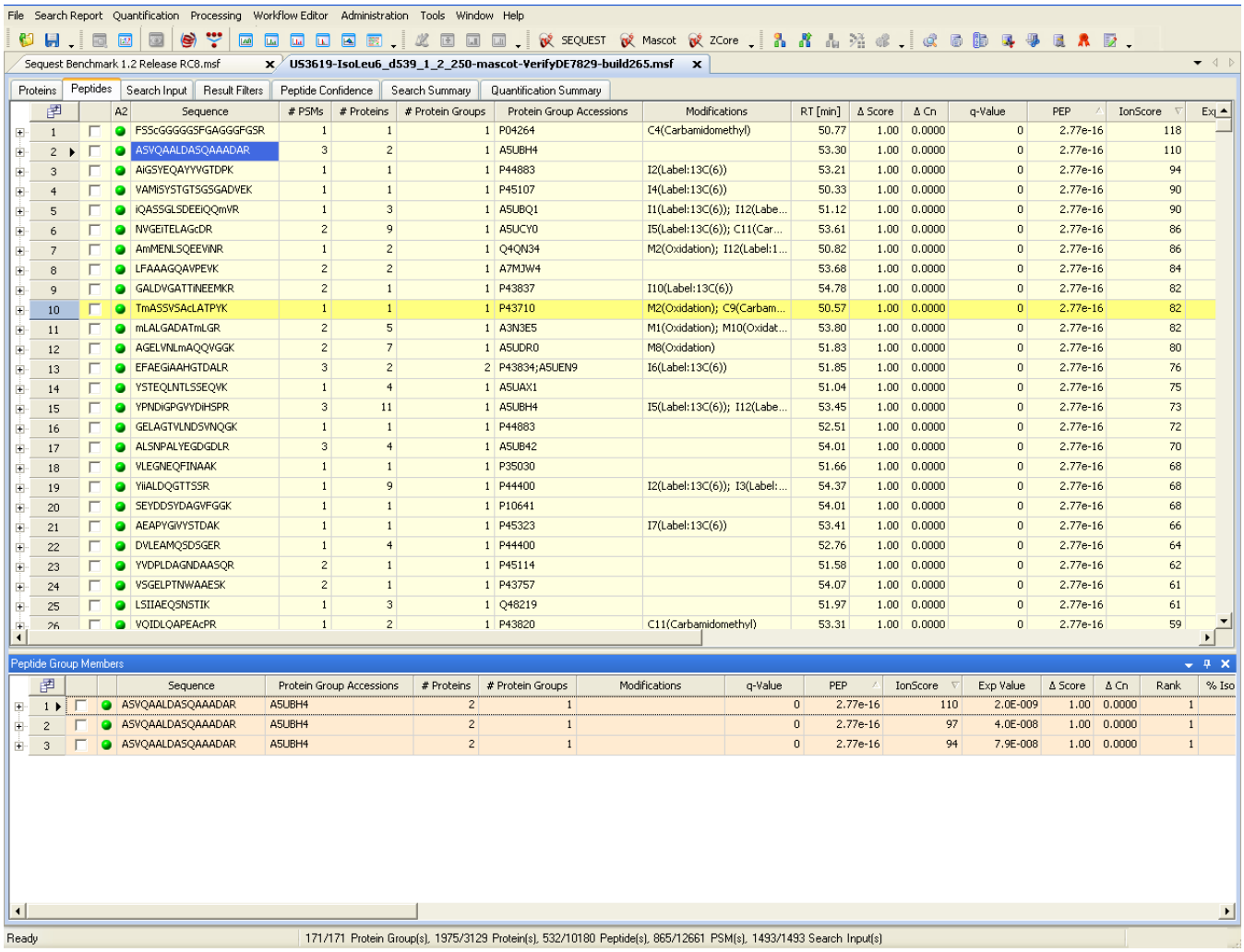
Figure 221. Workflow with Percolator attached to two different search nodes



4. Click on the **Percolator** node.
5. In the Maximum Delta Cn box in the parameters list, specify the Delta Cn value. For information on this parameter, see "Filtering Peptides by Delta Cn Value" on page 385.
6. In the Target FDR (Strict) box, set the target FDR for high-confidence peptide hits.
7. In the Target FDR (Relaxed) box, set the target FDR for medium-confidence peptide hits.
8. In the Validation Based On box, select either **q-Value** or **PEP** (posterior error probability) to assign to the target and decoy PSMs. For more information on these options, see "Percolator Node" on page 681.
9. Choose **Workflow Editor > Start Workflow**, or click the **Start Workflow** icon, .

When you open results processed with the Percolator node, each PSM and peptide group has two additional scores on the Peptides page, a q-value score and a posterior error probability (PEP) value, as shown in Figure 222.

Figure 222. PEP and q-Value columns on the Peptides page of results processed with Percolator



	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	RT [min]	Δ Score	Δ Cn	q-Value	PEP	IonScore
1		FSScGGGGGSFGAGGGFGSR	1	1	1	P04264	C4(Carbamidomethyl)	50.77	1.00	0.0000	0	2.77e-16	118
2		ASVQAALDASQAAADAR	3	2	1	A5UBH4		53.30	1.00	0.0000	0	2.77e-16	110
3		AiGSYEQAYYVGTDPK	1	1	1	P44883	I2(Label:13C(6))	53.21	1.00	0.0000	0	2.77e-16	94
4		VAMiSYSTGTSGSGADVEK	1	1	1	P45107	I4(Label:13C(6))	50.33	1.00	0.0000	0	2.77e-16	90
5		iQASSGLSDEEiQQmVR	1	3	1	A5UBQ1	I1(Label:13C(6)); I12(Labe...	51.12	1.00	0.0000	0	2.77e-16	90
6		NVGEiTELAGcDR	2	9	1	A5UCY0	I5(Label:13C(6)); C11(Car...	53.61	1.00	0.0000	0	2.77e-16	86
7		AmMENLSQEEViNR	1	2	1	Q4QN34	M2(Oxidation); I12(Label:1...	50.82	1.00	0.0000	0	2.77e-16	86
8		LFAAAGQAVPEVK	2	2	1	A7MJW4		53.68	1.00	0.0000	0	2.77e-16	84
9		GALDVGATTiNEEMKR	2	1	1	P43837	I10(Label:13C(6))	54.78	1.00	0.0000	0	2.77e-16	82
10		TmASSVSAcLATPYK	1	1	1	P43710	M2(Oxidation); C9(Carbam...	50.57	1.00	0.0000	0	2.77e-16	82
11		mLALGADATmLGR	2	5	1	A3N3E5	M1(Oxidation); M10(Oxidat...	53.80	1.00	0.0000	0	2.77e-16	82
12		AGELVNLmAQQVGGK	2	7	1	A5UDR0	M8(Oxidation)	51.83	1.00	0.0000	0	2.77e-16	80
13		EFAEGiAAHGTDALR	3	2	2	P43834;A5UEN9	I6(Label:13C(6))	51.85	1.00	0.0000	0	2.77e-16	76
14		YSTEQLNTLSSEQVK	1	4	1	A5UAX1		51.04	1.00	0.0000	0	2.77e-16	75
15		YPNDiGPGVYDiHSPR	3	11	1	A5UBH4	I5(Label:13C(6)); I12(Labe...	53.45	1.00	0.0000	0	2.77e-16	73
16		GELAGTVLNDSVNQGK	1	1	1	P44883		52.51	1.00	0.0000	0	2.77e-16	72
17		ALSNPALYEGDGDLR	3	4	1	A5UB42		54.01	1.00	0.0000	0	2.77e-16	70
18		VLEGNEQFINAAK	1	1	1	P35030		51.66	1.00	0.0000	0	2.77e-16	68
19		YiiALDQGTTSSR	1	9	1	P44400	I2(Label:13C(6)); I3(Label:...	54.37	1.00	0.0000	0	2.77e-16	68
20		SEYDDSYDAGVFGGK	1	1	1	P10641		54.01	1.00	0.0000	0	2.77e-16	68
21		AEAPYGiVYSTDAK	1	1	1	P45323	I7(Label:13C(6))	53.41	1.00	0.0000	0	2.77e-16	66
22		DVLEAMQSDSGER	1	4	1	P44400		52.76	1.00	0.0000	0	2.77e-16	64
23		YVDPLDAGNDAASQR	2	1	1	P45114		51.58	1.00	0.0000	0	2.77e-16	62
24		VSGELPTNWAAESK	2	1	1	P43757		54.07	1.00	0.0000	0	2.77e-16	61
25		LSIIAEQSNSTIK	1	3	1	Q48219		51.97	1.00	0.0000	0	2.77e-16	61
26		VQIDLQAPEAcPR	1	2	1	P43820	C11(Carbamidomethyl)	53.31	1.00	0.0000	0	2.77e-16	59

	Sequence	Protein Group Accessions	# Proteins	# Protein Groups	Modifications	q-Value	PEP	IonScore	Exp Value	Δ Score	Δ Cn	Rank	% Iso
1	ASVQAALDASQAAADAR	A5UBH4	2	1		0	2.77e-16	110	2.0E-009	1.00	0.0000	1	
2	ASVQAALDASQAAADAR	A5UBH4	2	1		0	2.77e-16	97	4.0E-008	1.00	0.0000	1	
3	ASVQAALDASQAAADAR	A5UBH4	2	1		0	2.77e-16	94	7.9E-008	1.00	0.0000	1	

Viewing the Results on the Peptide Confidence Page

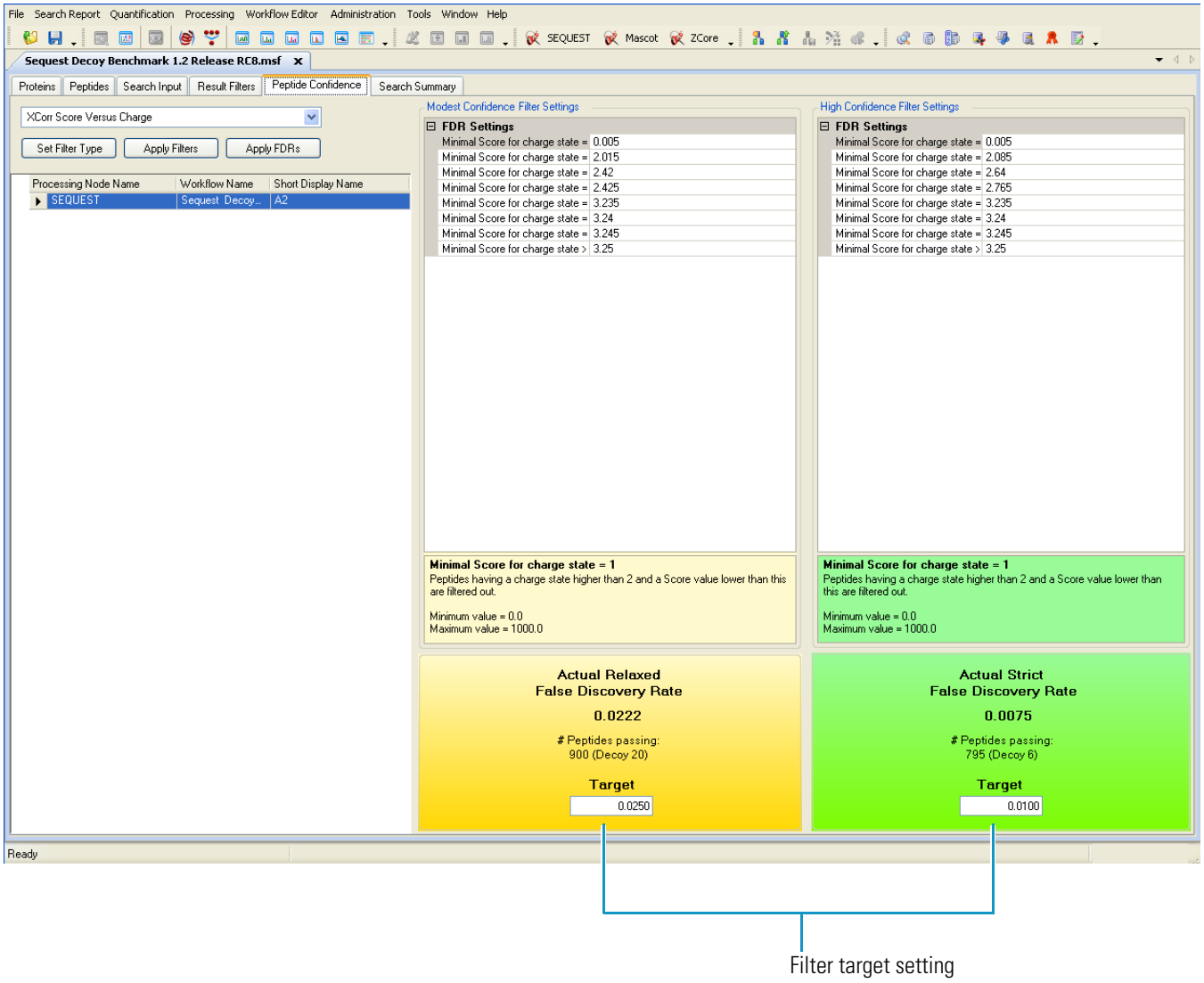
After the Proteome Discoverer application completes the search, open the results (MSF) file and view the decoy database search results on the Peptide Confidence page. This page shows the relaxed and strict false discovery rates with their corresponding filter settings listed above them.

❖ To display the Peptide Confidence page

- In an open report, click the **Peptide Confidence** tab.

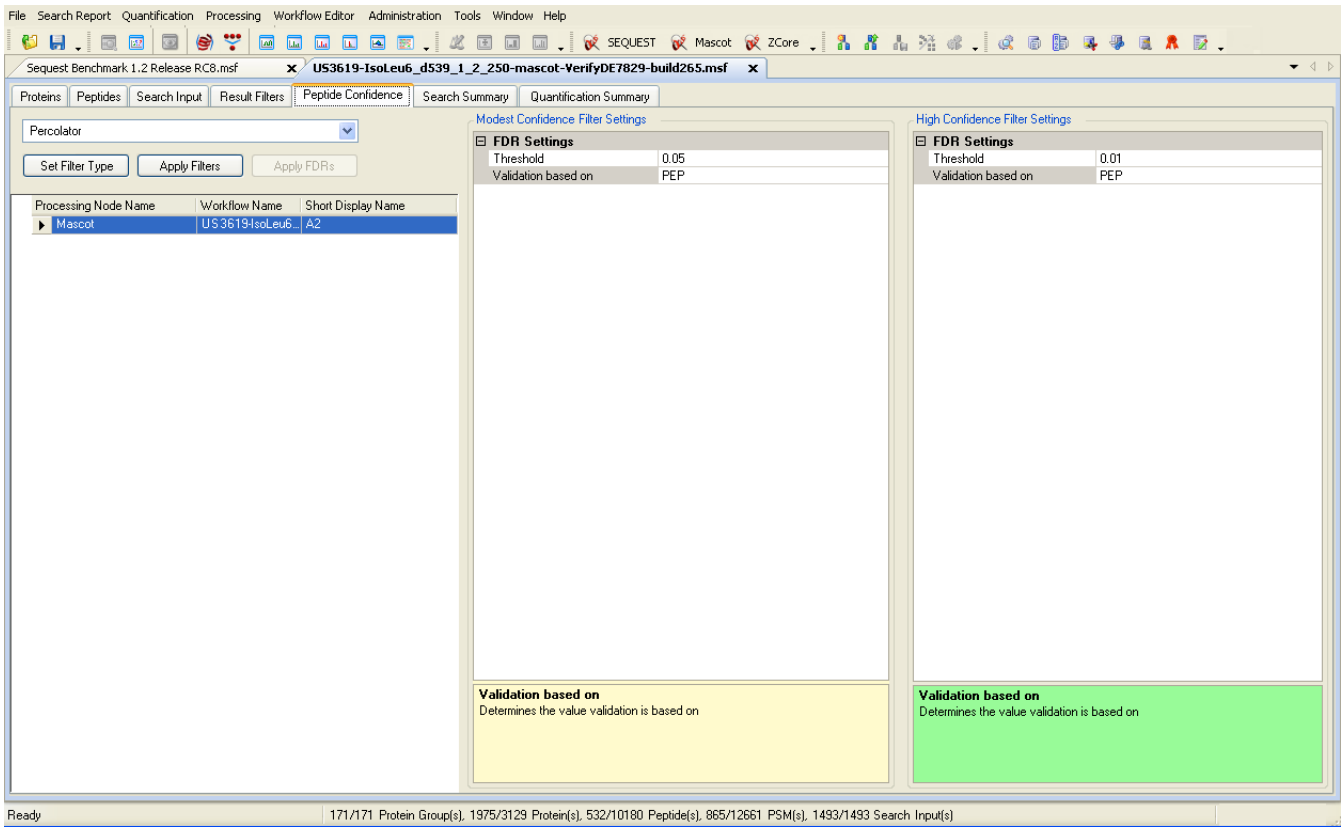
  The Peptide Confidence page of your search report appears, as shown in Figure 223. It filters out peptides to two predefined false discovery rates and sets the confidence levels for database searches.

Figure 223. Peptide Confidence page with the actual relaxed and strict false discovery rates



If you used the Percolator node in the workflow, you can set thresholds for the Percolator scores to separate PSMs of high confidence, medium confidence, and low confidence, as shown in Figure 224.

Figure 224. Setting thresholds for Percolator scores



In the box in the upper left of the Peptide confidence page, you can switch between validation based on Percolator and validation based on the calculation of target- and decoy-estimated FDRs from the search engine scores. This choice is always available, even if Percolator refused to process the data because it did not meet one of the requirements for the number of target and decoy matches.

Use the Peptide Confidence page to do the following:

- Set new filters and recalculate new false discovery rates based on these new filter criteria.
- Set new target false discovery rates and then recalculate new filter settings that, when applied, lead to false discovery rates no higher than the new target.

Note If you filter on peptide confidence during the loading of the report, all of the options on the Peptide Confidence page are unavailable because you can no longer adjust the settings.

Recalculating the False Discovery Rates

You can recalculate the false discovery rate on the Peptide Confidence page.

❖ To recalculate the false discovery rates

1. Open an MSF file, and click the **Peptide Confidence** tab.
2. In the Please select a filter list, select the approach for determining the peptide confidence. The available options are different for each search engine:
   - SEQUEST:
     - Peptide Score: Specifies that Peptide Score be used to calculate the FDR for determining peptide confidence.
     - (Default) Score versus Charge State: Specifies that Score Versus Charge State be used to calculate the FDR for determining peptide confidence.
     - Percolator: Specifies that Percolator be used to estimate peptide confidence based on the PSM features listed in Table 134.
   - Mascot:
     - Peptide Score: Specifies that the peptide score be used to calculate the FDR for determining peptide confidence.
     - (Default) Mascot Significance Threshold: Specifies that Mascot Significance Threshold be used to calculate the FDR for determining peptide confidence.
     - Percolator: Specifies that Percolator be used to estimate peptide confidence based on the PSM features listed in Table 134.
   - ZCore:
     - (Default) Peptide Score: Specifies that Peptide Score be used to calculate the FDR for determining peptide confidence.
     - Percolator: Specifies that Percolator be used to estimate peptide confidence based on the PSM features listed in Table 134.
3. Click **Set Filter Type** to apply the option that you selected in the Filter list to the settings in the Modest Confidence Filter Settings and the High Confidence Filter Settings panes.

Changing the Target Rate and Filter Settings

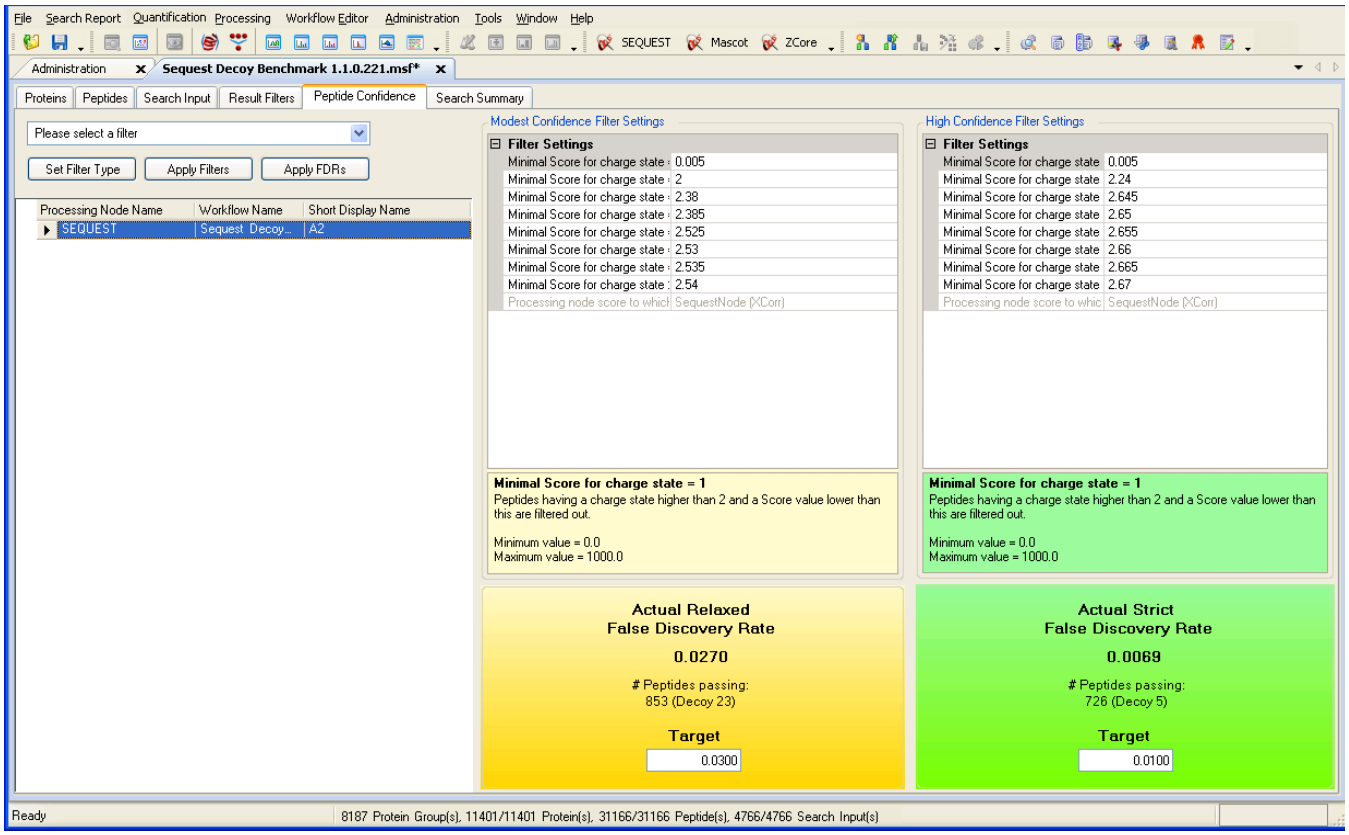
You can change the filter settings on the Peptide Confidence page by changing the target rate or changing the filter settings.

If you change the target rate or the filter settings, the application finds the actual relaxed or strict false discovery rate, or both, that come the closest to your target rate. It displays this number under Actual Relaxed False Discovery Rate or Actual Strict False Discovery Rate. It also displays the number of peptides and decoy peptides that pass the filters set in the Filter Settings area and changes the filter settings in the Filter Settings area.

Whether you change the target rate or the filter settings, the Proteome Discoverer application updates the peptide confidence indicators in the MSF report.

As an example, Figure 225 shows the results of entering a new target rate of 0.030 in the Target box of the Actual Relaxed False Discovery Rate area of the Peptide Confidence page shown in Figure 223 on page 424.

Figure 225. Results of new relaxed target rate



Go to the following sections:

- To change the target rate
- To change the filter settings
- To save the peptide confidence and FDR settings on the Results Filters page

❖ To change the target rate

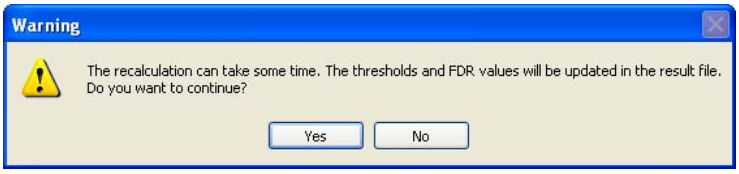
1. Change the value in the Target box of the Actual Relaxed False Discovery Rate area for medium confidence or the Actual Strict False Discovery Rate area for high confidence, or both.
2. Click **Apply FDRs**.

❖ To change the filter settings

1. Select the settings that you want to change in the Filter Settings area, and enter the new values. Minimal Score for Charge State specifies the charge state above which peptides are filtered out. Charge state values can range from 0 to 20.
2. Click **Apply Filters**.

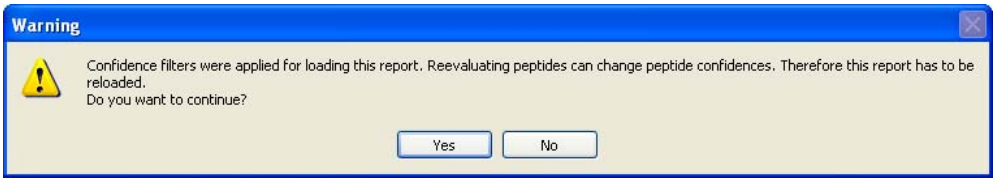
   The warning shown in Figure 226 now appears, unless you chose the Peptide Confidence filter in step 1.

   Figure 226. FDR recalculation message box for all filters except Peptide Confidence

   If you chose the Peptide Confidence filter in step 1, the message box shown in Figure 227 appears.

   Figure 227. FDR recalculation message box for Peptide Confidence filters

3. In either box, click **Yes**.

❖ To save the peptide confidence and FDR settings on the Results Filters page

- Choose **File > Save Report**.

Peptide Confidence Page Parameters

Table 67 describes the parameters on the Peptide Confidence page.

Table 67. Peptide Confidence page parameters (Sheet 1 of 3)

Parameters	Description
Filter box	Specifies the approach used to determine the peptide confidence. The options available depend on the search engine. • SEQUEST: – Peptide Score: Specifies that Peptide Score be used to calculate the FDR for determining peptide confidence. – (Default) Score versus Charge State: Specifies that Score Versus Charge State be used to calculate the FDR for determining peptide confidence. – Percolator: Specifies that Percolator be used to estimate peptide confidence based on the PSM features listed in Table 134. • Mascot: – Peptide Score: Specifies that Peptide Score be used to calculate the FDR for determining peptide confidence. – (Default) Mascot Significance Threshold: Specifies that Mascot Significance Threshold be used to calculate the FDR for determining peptide confidence. – Percolator: Specifies that Percolator be used to estimate peptide confidence based on the PS M features listed in Table 134. • ZCore: – Peptide Score: Specifies that Peptide Score be used to calculate the FDR for determining peptide confidence. – Percolator: Specifies that Percolator be used to estimate peptide confidence based on the PSM features listed in Table 134.
Set Filter Type	Reconfigures the filter type in the Modest Confidence Filter Settings and the High Confidence Filter Settings panes according to the option that you selected in the Please select a filter box.
Apply Filters	Applies your edited filter settings to the MSF report to change the report's peptide confidence indicators and updates the target FDR to equal the actual FDR.

Table 67. Peptide Confidence page parameters (Sheet 2 of 3)

Parameters	Description
Apply FDRs	Calculates the FDR on the basis of the target setting and updates the peptide confidences accordingly.
Processing Node Names	Specifies the name of the search engine used in the decoy database search.
Workflow Name	Specifies the name of the workflow used in the decoy database search.
Short Display Name	Displays an abbreviated name consisting of a letter and a number. • The letter corresponds to the numerical order of the opened report (for example, the first report opened is A, the second report opened is B, and so forth). • The number is the number assigned to the processing node in the Workflow Editor and shown on the Search Summary page.
Modest Confidence Filter Settings	Specifies the target rate for medium confidence. The default is specified by the Target FDR (Relaxed) parameter in the Workflow Editor or in the search engine wizard dialog box activated by the Processing > Start *Wizard_name* command.
Filter Settings	Shows the scoring criteria for determining the peptide confidence. For Peptide Score: • Score Threshold: Specifies a score below which peptides are classified as low confidence. Score values can be from 0 to 1000. • Processing node score that the filter is applied to: Displays the setting that you selected for the Processing node score name option of the Peptide Score filter on the Result Filters page. This parameter is read-only. For Score versus Charge State: • Minimal Score for charge state = *x*: Specifies the score above which peptides are classified as high confidence. Charge state values can range from 0 to 20. • Processing node score that the filter is applied to: Displays the setting that you selected for the Processing node score name option of the Score versus Charge State filter on the Result Filters page. This parameter is read-only.

Table 67. Peptide Confidence page parameters (Sheet 3 of 3)

Parameters	Description
Actual Relaxed False Discovery Rate	Displays the relaxed false discovery rate that is the closest to (but not greater than) your target rate.
# Peptides passing	Displays the number of peptides and decoy peptides that passed the filters specified in the Filter Settings area.
Target	Specifies the target for medium-confidence peptides.
High Confidence Filter Settings	Specifies the target rate for high confidence. The default is specified by the Target FDR (Strict) parameter in the Workflow Editor or in the search engine wizard dialog box activated by the Processing > Start *Wizard_name* command.
Filter Settings	Sets the criteria for filtering out peptides. For Peptide Score: • Score Threshold: Specifies a score below which peptides are classified as low-confidence. Score values can be from 0 to 1000. • Processing node score that the filter is applied to: Displays the setting that you selected for the Processing node score name option of the Peptide Score filter on the Result Filters page. This parameter is read-only. For Score versus Charge State: • Minimal Score for charge state = *x*: Specifies the score above which peptides are classified as high-confidence. Charge state values can range from 0 to 20. • Processing node score that the filter is applied to: Displays the setting that you selected for the Processing node score name option of the Score versus Charge State filter on the Result Filters page. This parameter is read-only.
Actual Strict False Discovery Rate	Displays the strict false discovery rate that is the closest to (but not greater than) your target rate.
# Peptides passing	Displays the number of peptides and decoy peptides that passed the filters specified in the Filter Settings area.
Target	Specifies the target rate for high-confidence peptides.

Protein Annotation

This chapter explains how the Proteome Discoverer application retrieves annotation information from ProteinCenter, including GO (Gene Ontology) annotations, Pfam (Protein Families) annotations, and information about post-translational modifications (PTMs) from UniProt.

Contents

- ProteinCenter
- Gene Ontology (GO) Annotation
- Pfam Annotation
- Configuring Protein Annotation
- Creating a Protein Annotation Workflow
- Displaying the Annotated Protein Results
- Reannotating MSF Files
- Uploading Results to ProteinCenter
- Accessing ProteinCards
- ProteinCard Parameters
- GO Slim Categories

ProteinCenter

ProteinCenter is a Web-based application that you can use to download biologically enriched annotation information for a single protein, such as molecular functions, cellular components, and biological processes, from the GO and Pfam databases. The Proteome Discoverer application gives you access to ProteinCenter in two ways:

- The Annotation node used in a search workflow retrieves GO, Pfam, and Uniprot database information from ProteinCenter and stores it in the Proteome Discoverer results files. This information is displayed in the Molecular Functions, Cellular Components, and Biological Processes columns on the Proteins page of the MSF file. For information on setting up an Annotation workflow to achieve these results, see "Configuring Protein Annotation" on page 435 and "Creating a Protein Annotation Workflow" on page 437.
- The ProteinCard available for each protein displays the annotation data available in ProteinCenter and displays it on a page of the Protein Identification Details dialog box (see "Accessing ProteinCards" on page 449). You can display this information for the following proteins:
  - Proteins on the Proteins page of the MSF file
  - Proteins associated with identified peptides
  - Proteins shown in the Protein Group Members view

  You can access the ProteinCard for each protein by double-clicking its row in the MSF report or clicking its row and choosing Search Report > Show Protein ID Details and then clicking the ProteinCard tab of the Protein Identification Details dialog box. The ProteinCard itself is split into separate tabs representing different aspects of that protein: General, Keys, Features, Molecular Functions, Cellular Components, Biological Processes, Diseases, and External Links. You can display a ProteinCard for every identified protein whose accession is tracked in ProteinCenter. For information on ProteinCard, see "Accessing ProteinCards" on page 449 and "ProteinCard Parameters" on page 450.

  You can also upload protein results directly from the Proteome Discoverer application to ProteinCenter. For information, see "Uploading Results to ProteinCenter" on page 446.

Gene Ontology (GO) Annotation

The Gene Ontology (GO) database is a collaborative effort, incorporating community input from database and genome annotation groups to address the need for consistent descriptions of gene products in different databases. The GO project has developed three structured, controlled vocabularies (ontologies) that describe gene products in a species-independent manner.

biological processes

cellular components

molecular functions

Each gene ontology is divided into categories and subcategories called GO terms, which define the protein in more specific terms. For example, chloroplast, a term in the cellular component ontology, is subdivided as follows:

chloroplast

[p] *chloroplast* envelope

[p] *chloroplast* membrane

[i] *chloroplast* inner membrane

[i] *chloroplast* outer membrane

You can obtain more information on the GO Ontology Web site at www.geneontology.org/.

Pfam Annotation

In addition to GO annotations, you can also retrieve from ProteinCenter Pfam annotations from the Pfam database (//pfam.sanger.ac.uk). These are annotations of protein families, which are proteins with similar sequences and similar biological functions. A special sequence comparison algorithm called the Hidden Markov Model groups proteins into the families by comparing the sequences. Each family has its own ID number that starts with Pf … . The Proteins page of the MSF file displays this number in the Pfam IDs column. You can use the Pfam identification number to go to the Pfam database to obtain more details about the protein family. You can also activate the ProteinCard for each protein by double-clicking on the Pfam identification number.

The Pfam annotation system is an alternative to GO annotations. You might want to use the Pfam system to filter your proteins when you want the results to be traceable, scored, and uniformly grouped. You might also consider its computationally based data more reliable. However, it might be easier to use the hierarchy and grouping of the GO system to help you interpret results.

Table 68 compares the features of the GO and Pfam databases.

Table 68. Comparison of GO and Pfam features

GO features	Pfam features
Proteins grouped in biologically meaningful categories	Proteins grouped by similarity
Deep hierarchical order of terms	Few hierarchies
Data input by experts with different confidence levels and differing opinions	Computational data input with no human influence or expert knowledge

Configuring Protein Annotation

Before you can start a search that includes protein annotation in the results or display ProteinCards for proteins, you must configure the Proteome Discoverer application for protein annotation.

❖ To configure the application for protein annotation

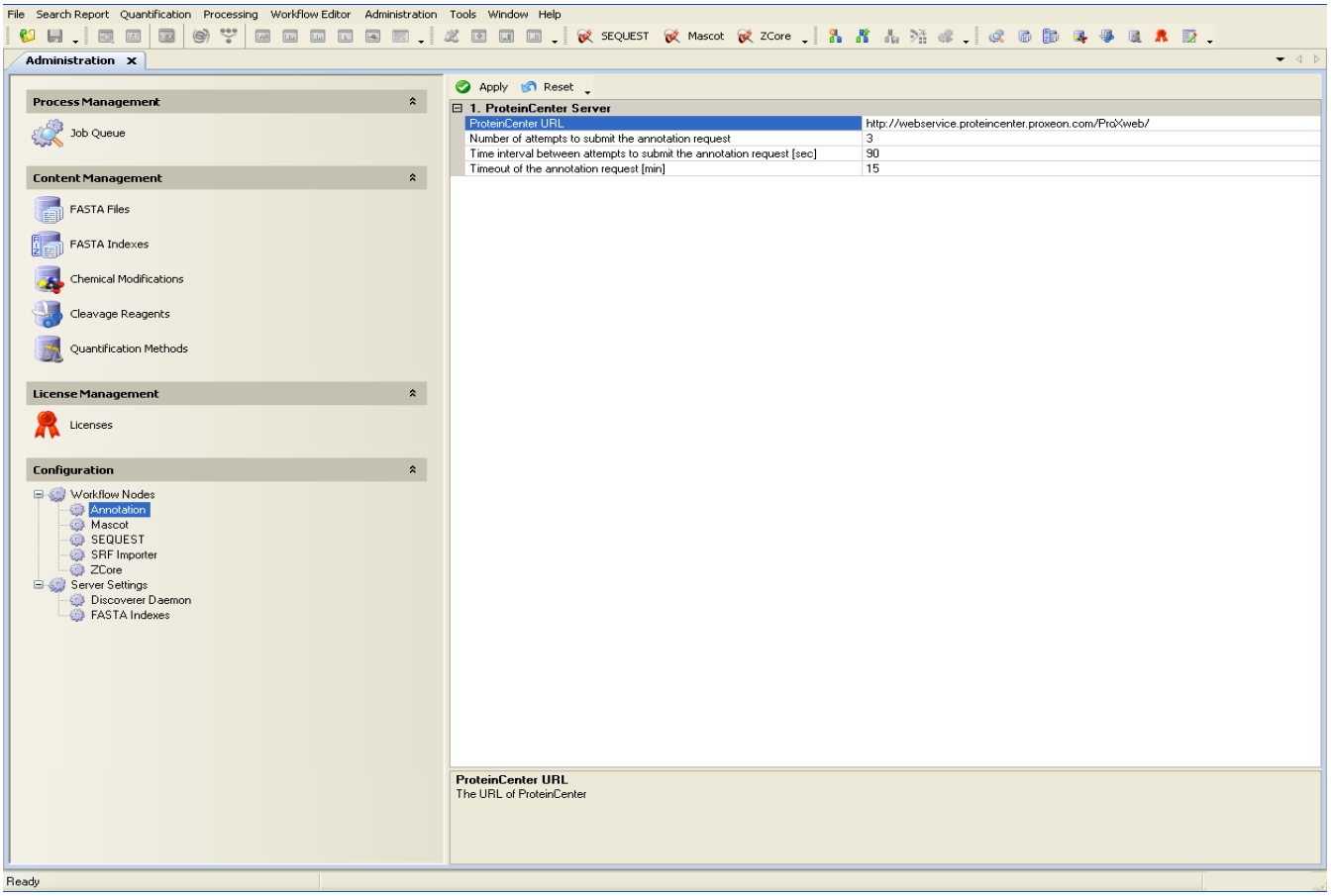
1. Choose **Administration > Configuration** or click the **Edit Configuration** icon,  .

   The Administration page changes to the Configuration view.

2. Under Workflow Nodes in the Configuration section of the left pane, click **Annotation**, if it is not already selected.

   The Annotation view appears, as shown in Figure 228.

Figure 228. Annotation view



3. In the ProteinCenter URL box, type the path and name of the ProteinCenter Web server.

   Thermo Fisher Scientific gives you this URL, a user name, and a password when you subscribe to ProteinCenter.

   Changes in the URL take effect after you restart the Proteome Discoverer application. If you entered an incorrect URL, the ProteinCard tab of the Protein Identification Details dialog box displays an error message.

4. In the Number of Attempts to Submit the Annotation Request box, specify the number of times that the Proteome Discoverer application should try to obtain the requested annotations if the ProteinCenter Web service issues an error.

   The default is 3.

5. In the Time Interval Between Attempts to Submit the Annotation Request [sec] box, specify the amount of time, in seconds, that the Proteome Discoverer application should wait between tries to obtain the requested annotations if the ProteinCenter Web service issues an error.

   The default is 90 seconds.

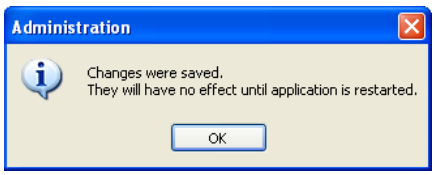
6. In the Timeout of the Annotation Request [min] box, specify the amount of time, in minutes, that the Proteome Discoverer application should continue to try to access the ProteinCenter Web service.

   The default is 15 minutes.

7. If you changed any settings, click Apply.

   The message box shown in Figure 229 appears:

   Figure 229. Administration message box



8. Click **OK**.

   Note Click Reset to return to the default values.

9. Restart your machine.

Creating a Protein Annotation Workflow

You can retrieve annotations of all identified proteins from ProteinCenter by using the Annotation node in a workflow. This node can retrieve Gene Ontology (GO) and GO slim annotations, protein family (Pfam) annotations, and annotations of modifications documented in the UniProt database. It stores this information in the Proteome Discoverer results files. GO Slim annotations are displayed in the Molecular Function, Cellular Component, and Biological Process columns in the MSF file. The Proteome Discoverer application retrieves the annotation data after all the search nodes have finished processing.

The data in ProteinCenter is updated biweekly.

❖ To create an annotation workflow

1. Choose **Workflow Editor > New Workflow**.
2. Set up your workflow by following the instructions in "Starting a New Search by Using the Workflow Editor" on page 89.
3. In the Annotation area of the Workflow Nodes pane, select the **Annotation** node and drag it to the Workspace pane.

   The Annotation node automatically connects to the other nodes in the workflow.
4. (Optional) After you join all your chosen nodes, align them by choosing **Workflow Editor > Auto Layout** or clicking the **Auto Layout** icon () or right-clicking a node and choosing **Auto Layout** from the shortcut menu.
5. (Optional) You can renumber the workflow nodes in the workflow in consecutive order by choosing **Workflow Editor > Auto Number**.
6. Choose **Workflow Editor > Start Workflow** or click the **Start Workflow** icon,  .

Displaying the Annotated Protein Results

The Proteome Discoverer application retrieves GO, Pfam, and UniProt PTM annotation data from ProteinCenter when it finishes processing all search nodes. You can display the annotated protein results in the MSF file. For GO annotations, the application can filter the list of identified proteins by selected Go Slim categories.

Note The Proteome Discoverer application cannot retrieve annotations from searches conducted in the UniRef FASTA database because of the prefix appended to the accession number.

- To display the GO protein annotation results
- To filter the identified proteins by GO Slim categories
- To display Protein Family (Pfam) annotation results

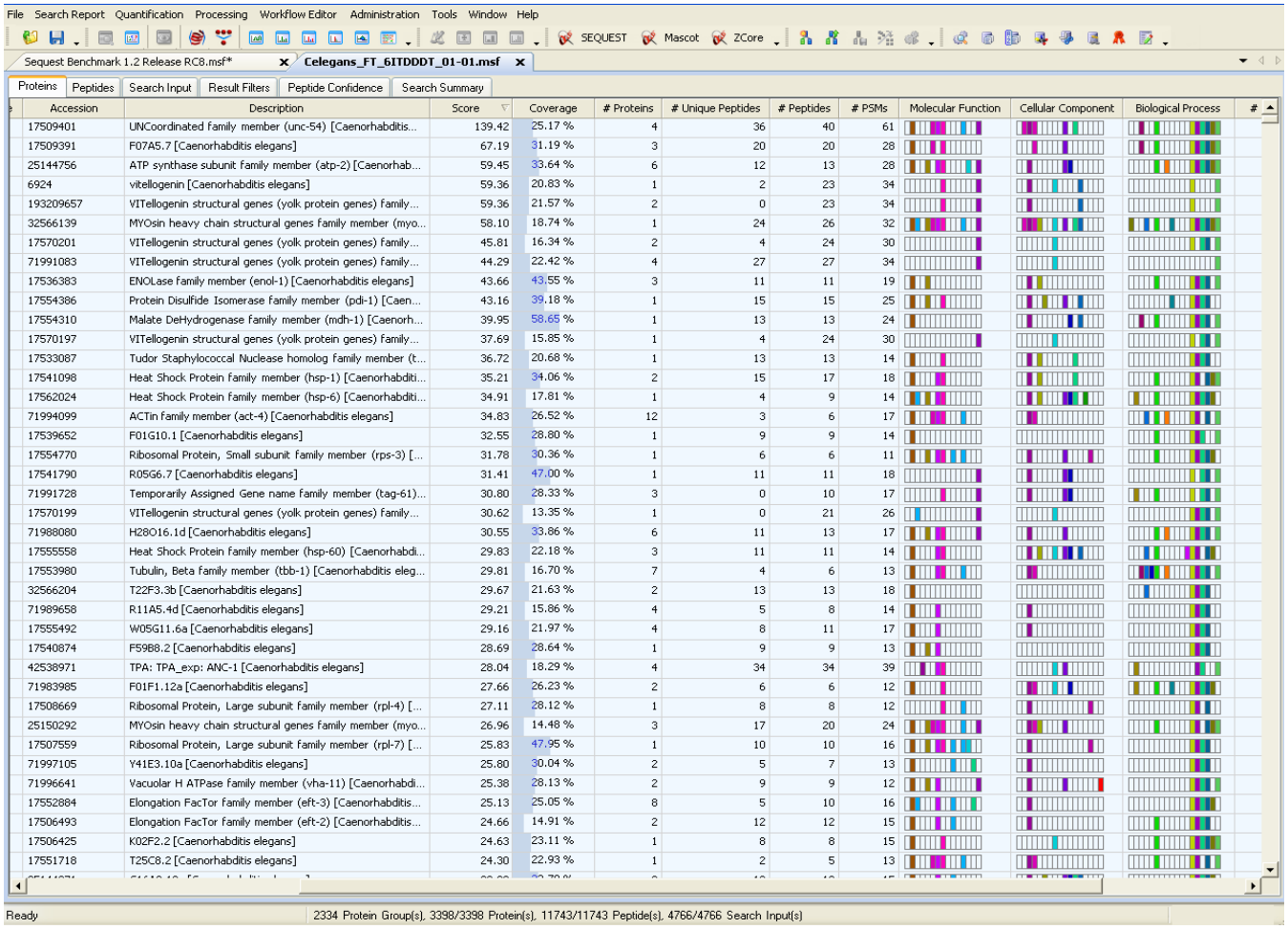
❖ To display the GO protein annotation results

1. Open the generated MSF file by following the instructions in "Opening MSF Files" on page 215.
2. In the Column Chooser dialog box of the proteins page, select the Molecular Function, Cellular Component, and Biological Processes columns.

   For information on the Column Chooser dialog box, see "Selecting the Columns to Display" on page 223.

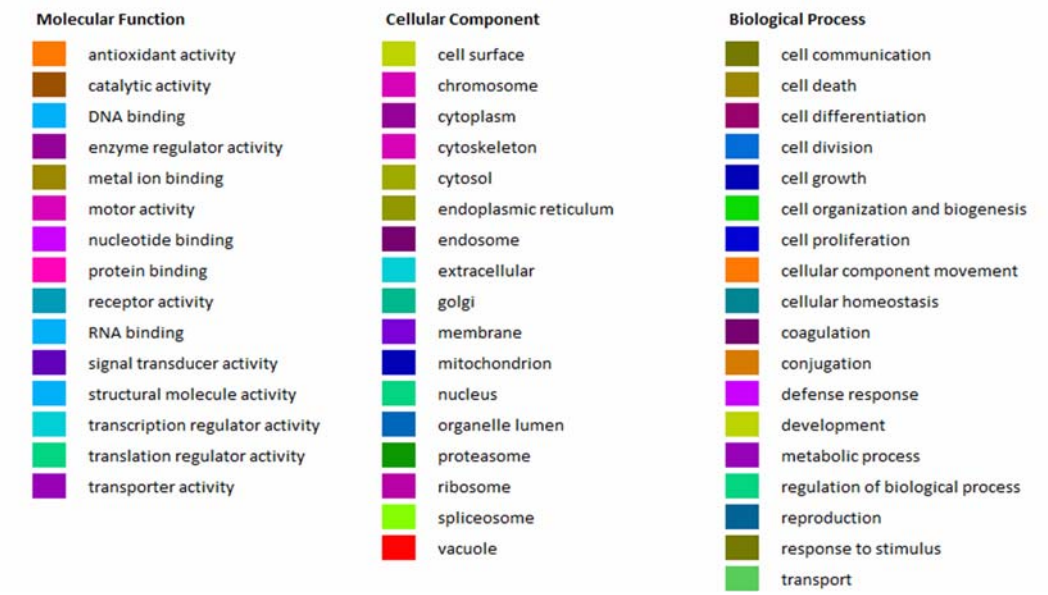
   The Proteome Discoverer application displays the results on the Proteins page of the MSF report as colored boxes similar to those shown in ProteinCenter. Figure 230 gives an example. If it does not find the requested protein in ProteinCenter, it displays a "protein not found" message in the annotation columns. If the annotation retrieval failed because of issues with the Web request, it displays an error message in the annotation columns.

Figure 230. GO Slim category boxes for the protein groups shown in the results of an annotation search



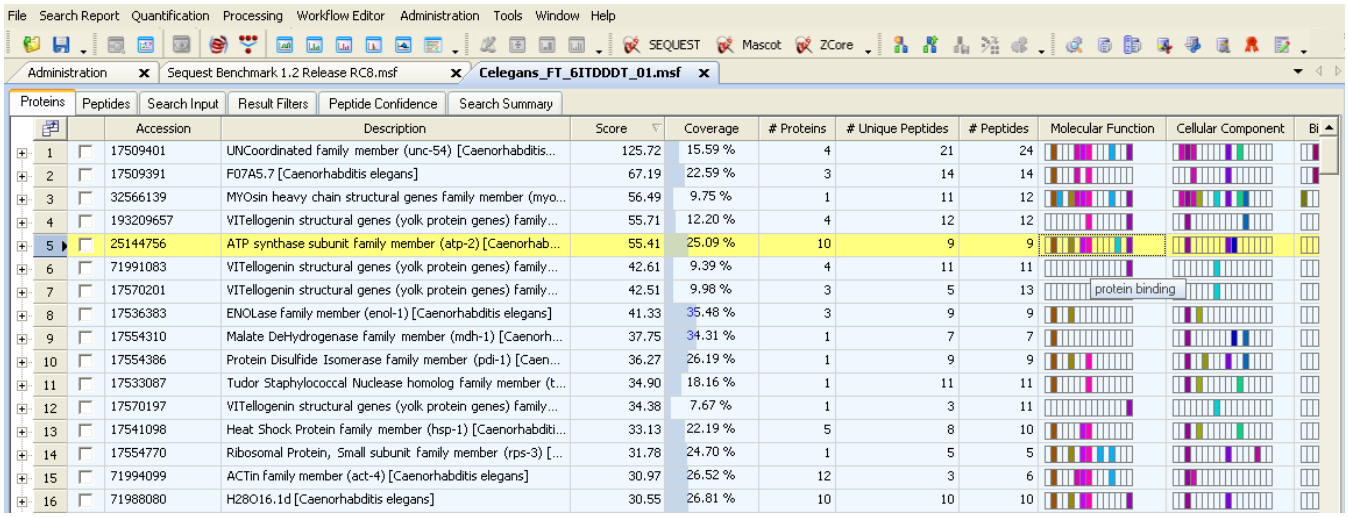
Each aspect of the annotation (biological processes, cellular components, and molecular functions) is represented in a separate column. Each box represents a GO Slim category, which is a selected subset of the Gene Ontology annotations. If the protein annotation is included in one of these subsets, the corresponding box is highlighted by a color specific to this GO Slim category. Figure 231 shows the meaning of the GO Slim category colors.

Figure 231. GO Slim category colors



If you move the mouse over the GO Slim category boxes, a ToolTip showing the category name appears, as shown in the Molecular Function column in Figure 232.

Figure 232. ToolTip identifying the annotation category



	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	Molecular Function	Cellular Component
1	17509401	UNCoordinated family member (unc-54) [Caenorhabditis...	125.72	15.59 %	4	21	24		
2	17509391	F07A5.7 [Caenorhabditis elegans]	67.19	22.59 %	3	14	14		
3	32566139	MYOsin heavy chain structural genes family member (myo...	56.49	9.75 %	1	11	12		
4	193209657	VITellogenin structural genes (yolk protein genes) family...	55.71	12.20 %	4	12	12		
5	25144756	ATP synthase subunit family member (atp-2) [Caenorhab...	55.41	25.09 %	10	9	9		
6	71991083	VITellogenin structural genes (yolk protein genes) family...	42.61	9.39 %	4	11	11		
7	17570201	VITellogenin structural genes (yolk protein genes) family...	42.51	9.98 %	3	5	13		
8	17536383	ENOLase family member (enol-1) [Caenorhabditis elegans]	41.33	35.48 %	3	9	9		
9	17554310	Malate DeHydrogenase family member (mdh-1) [Caenorh...	37.75	34.31 %	1	7	7		
10	17554386	Protein Disulfide Isomerase family member (pdi-1) [Caen...	36.27	26.19 %	1	9	9		
11	17533087	Tudor Staphylococcal Nuclease homolog family member (t...	34.90	18.16 %	1	11	11		
12	17570197	VITellogenin structural genes (yolk protein genes) family...	34.38	7.67 %	1	3	11		
13	17541098	Heat Shock Protein family member (hsp-1) [Caenorhabditi...	33.13	22.19 %	5	8	10		
14	17554770	Ribosomal Protein, Small subunit family member (rps-3) [...	31.78	24.70 %	1	5	5		
15	71994099	ACTin family member (act-4) [Caenorhabditis elegans]	30.97	26.52 %	12	3	6		
16	71988080	H28O16.1d [Caenorhabditis elegans]	30.55	26.81 %	10	10	10		

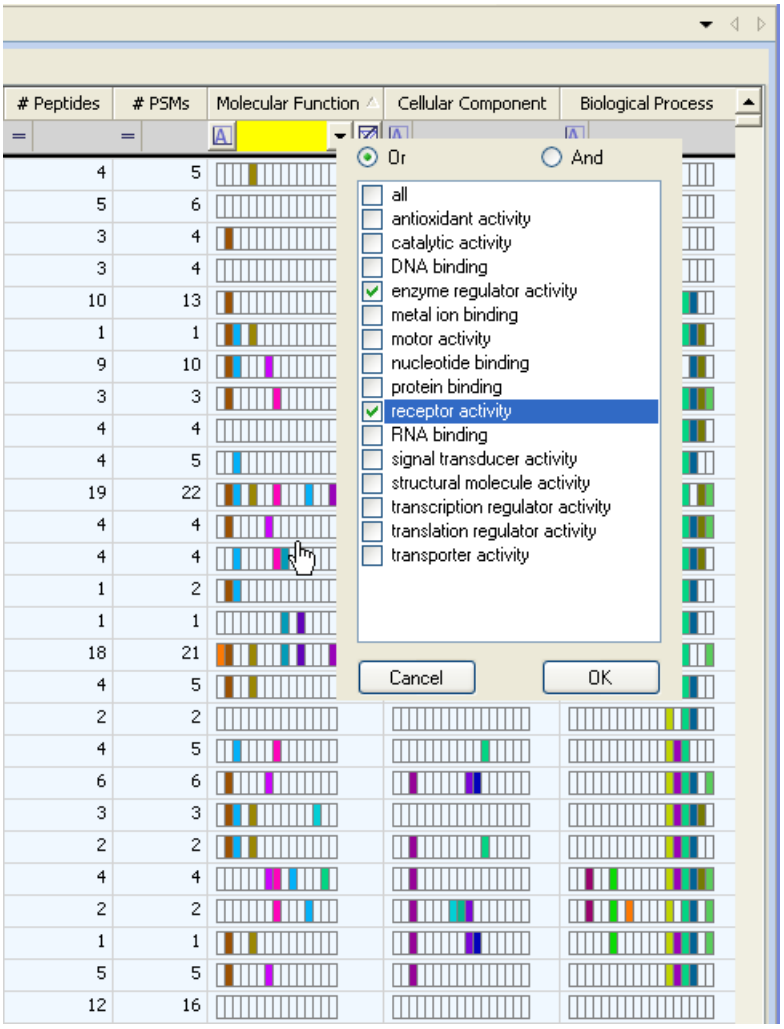
In multiconsensus reports, the protein information is displayed for the master protein of a protein group.

❖ To filter the identified proteins by GO Slim categories

1. In the MSF report, right-click the Proteins page and choose **Enable Row Filters**.
2. Click in the filter row that appears beneath the column headers in one of the GO columns, for example, Molecular Function.
3. Click the down arrow in this row.

   A dialog box appears that lists the GO Slim categories that you can filter by, as shown in Figure 233.

   Figure 233. Filtering by GO Slim category

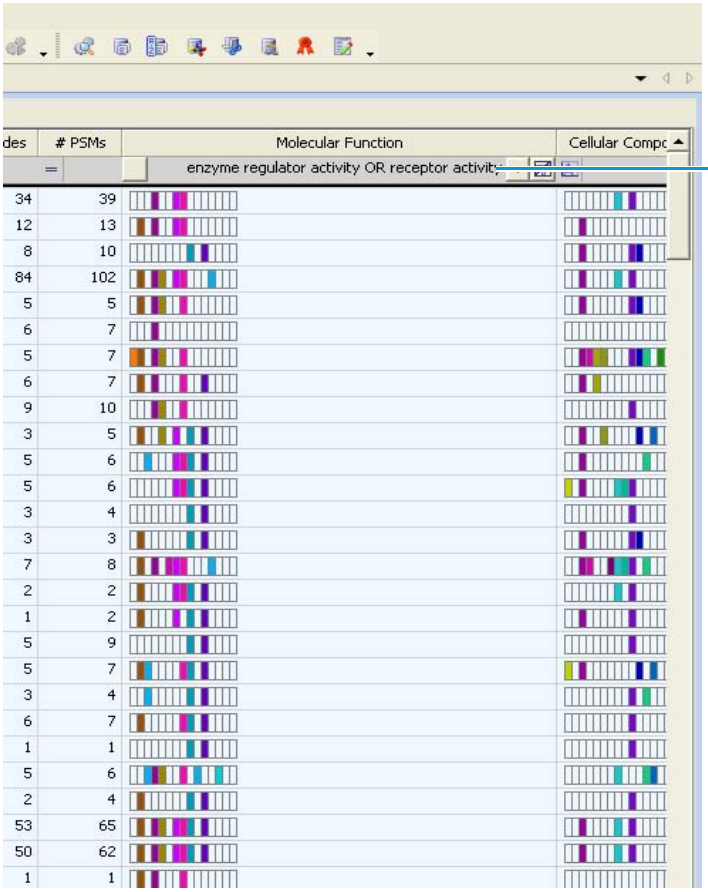


4. Select one or more of the GO Slim categories.
5. If you selected more than one GO Slim category, select the logical **And** option at the top of the dialog box to indicate that the Proteome Discoverer application should filter by the combined categories, or select the logical **Or** option to indicate that it should filter by only one category.

6. Click **OK**.

   The Proteome Discoverer application displays the identified proteins belonging to the selected categories. The names of the categories selected appear in the filter row when you expand the width of the column, as shown in Figure 234.

Figure 234. List of proteins filtered by Go Slim category



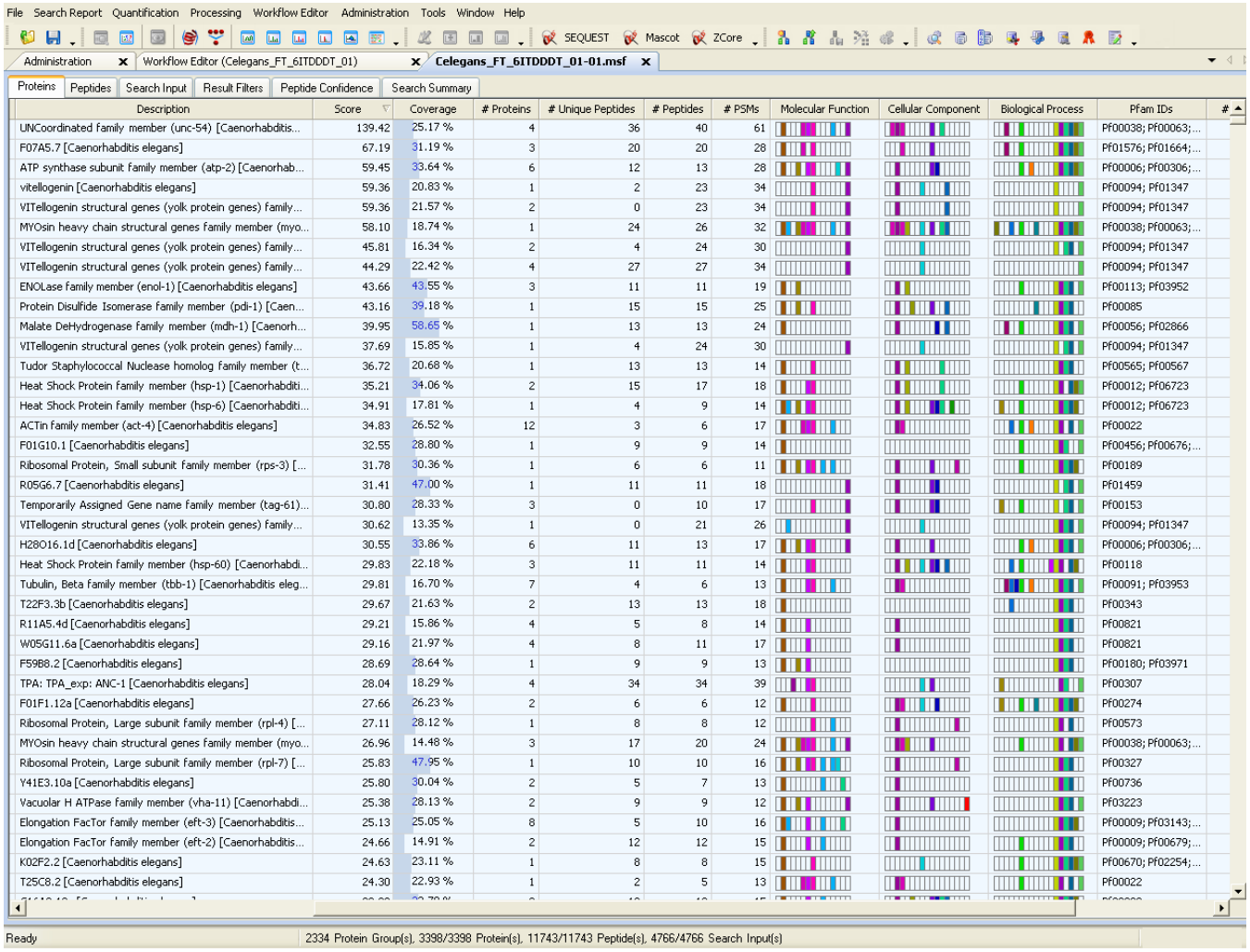
❖ To display Protein Family (Pfam) annotation results

1. Open the MSF file by following the instructions in "Opening MSF Files" on page 215.
2. In the Column Chooser dialog box of the proteins page, select the Pfam IDs column.

   For information on the Column Chooser dialog box, see "Selecting the Columns to Display" on page 223.

   Figure 235 shows the Pfam annotations in the Pfam IDs column.

Figure 235. Pfam annotations for the protein groups shown in the results of an annotation search



Reannotating MSF Files

You can use the Re-Annotation node in the Workflow Editor or the batch processing function in Discoverer Daemon to update existing annotations or annotate existing MSF files that do not yet include annotations.

Use the Re-Annotation node in the Workflow Editor to reannotate a single file. The Re-Annotation node must be the only node in a workflow. It takes an existing MSF file as input, retrieves up-to-date annotations for the proteins contained in the MSF file, and stores them in the same MSF file.

Note If you used a previous version of the Proteome Discoverer application to create the MSF file to reannotate, the application first updates the file to comply with the current results file schema.

Use the batch processing function in Discoverer Daemon to reannotate multiple files.

- To reannotate an MSF file in the Workflow Editor
- To reannotate an MSF file in Proteome Discoverer Daemon

❖ To reannotate an MSF file in the Workflow Editor

1. Choose **Workflow Editor > New Workflow**.
2. In the Annotation area of the Workflow Nodes pane, select the **Re-Annotation** node and drag it to the Workspace pane.
3. Click the MSF File Path box, and click the browse button (...) to open the Select Analysis File dialog box.
4. Browse to the MSF file to save the new annotations in, or type the path and name of the file in the File Name box, and click **Open**.

   The name of the MSF file appears in the Name box in the Workflow Editor.
5. Choose **Workflow Editor > Start Workflow** or click the **Start Workflow** icon,  .

The Proteome Discoverer application submits the workflow to standard workflow processing and displays the reannotation progress in the job queue.

Note If you created the MSF file that you want to reannotate with a previous version of the Proteome Discoverer application, the application updates the file first to comply with the current result file schema.

❖ To reannotate an MSF file in Proteome Discoverer Daemon

1. Create a reannotation workflow in the Workflow Editor according to the instructions in "To reannotate an MSF file in the Workflow Editor" on page 444.
2. Save the workflow as a new workflow template:
   a. Choose **Workflow Editor > Save As Template**.
   b. In the Save Processing Workflow Template dialog box, type the name of the template in the Template Name box.
   c. Give a brief description of the template in the Template Description box.
   d. Click **Save**.

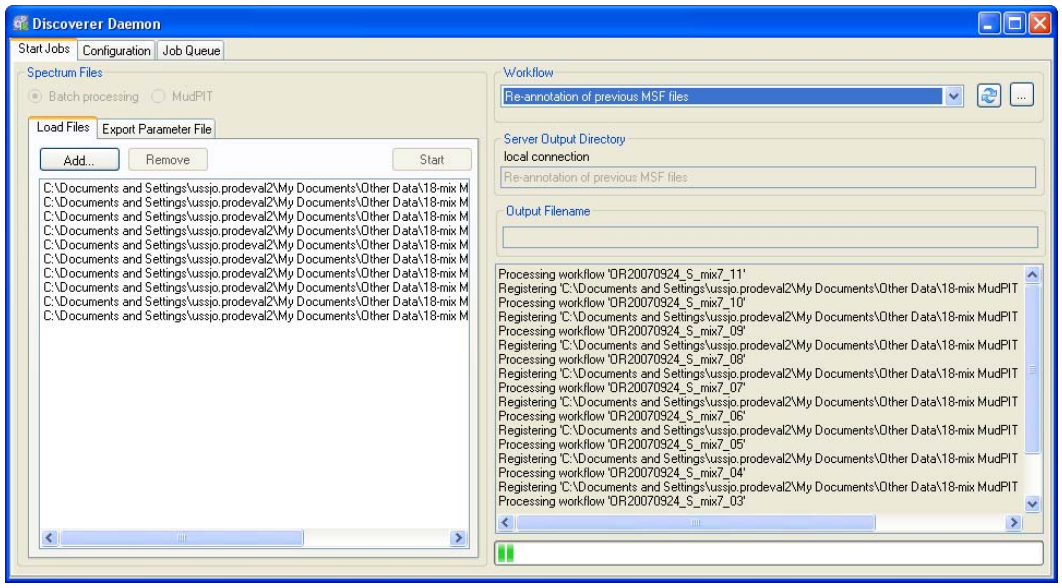
   This newly created workflow template is now available in Discoverer Daemon.
3. To start Discoverer Daemon, follow the instructions in "Starting the Proteome Discoverer Daemon Application in a Window" on page 121.
4. To select the server, follow the instructions in "Selecting the Server" on page 122.
5. Click the **Start Jobs** tab.

6. Click the **Load Files** tab, if it is not already selected.
7. Click **Add**.
8. In the Open dialog box, browse to the MSF file that you want to save the new annotations in, or type the name of the file in the File Name box, and click **Open**.
9. Repeat step 8 to add the names of multiple MSF files to reannotate.
10. In the Spectrum Files area, click **Batch Processing**.
11. From the menu in the Workflow box, select the reannotation workflow template that you saved in the Workflow Editor.
12. Start the batch processing:
    - If you are connected to an instance of the Proteome Discoverer application running on the same computer, click the **Start** button in Discoverer Daemon.
    - If you are connected to an instance of the Proteome Discoverer application running on a remote machine, specify in the Server Output Directory box the name of the folder where you want the original output files placed on the server, and then click the **Start** button.

      By default, the Proteome Discoverer Daemon application places this folder in the c:\Documents and Settings\All Users\...\DiscovererDaemon\SpectrumFiles\ directory. You can specify a different folder by choosing Administration > Configuration in the Proteome Discoverer application, clicking Discoverer Daemon in the Server Settings section, and browsing to the location in the New Directory box.

Figure 236 shows MSF files being processing in batch mode in Discoverer Daemon.

Figure 236. Reannotating MSF files in batch mode in Discoverer Daemon



For more information about processing files with Discoverer Daemon, see "Using the Proteome Discoverer Daemon Utility" on page 121.

Uploading Results to ProteinCenter

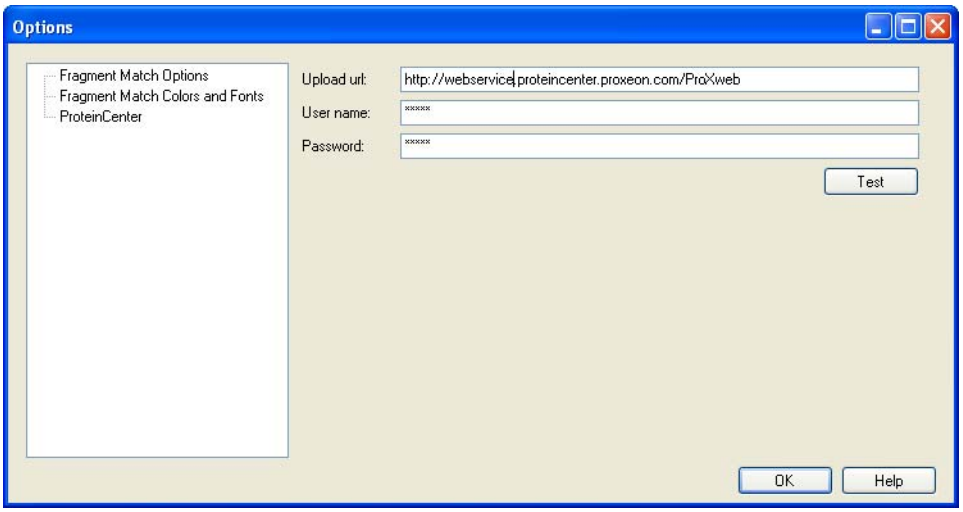
If you have a user account on a ProteinCenter server, you can upload search results directly from the Proteome Discoverer application to ProteinCenter.

❖ To upload search results to ProteinCenter

1. Choose **Tools > Options**.
2. In the Options dialog box, click **ProteinCenter**.

   The ProteinCenter page opens, as shown in Figure 237.

   Figure 237. ProteinCenter page of the Options dialog box



   a. In the URL box, type the URL of the ProteinCenter server to use.
   b. In the User Name box, type the user name of your ProteinCenter user account.
   c. In the Password box, type the password of your ProteinCenter user account.
   d. Click **OK**.

   A message box appears with the following message:

   Settings of Protein Center changed. Do you want to save your changes?

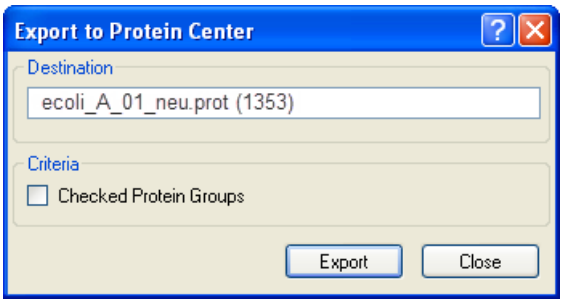
3. Click **Yes**.
4. Open an MSF file in the Proteome Discoverer application. See "Opening MSF Files" on page 215.

5. Choose **Tools > Export to ProteinCenter**.

   The Export to ProteinCenter dialog box opens.

6. In the Destination box, specify the name of the data set to upload to ProteinCenter, as shown in Figure 238.

   Figure 238. Export to ProteinCenter dialog box



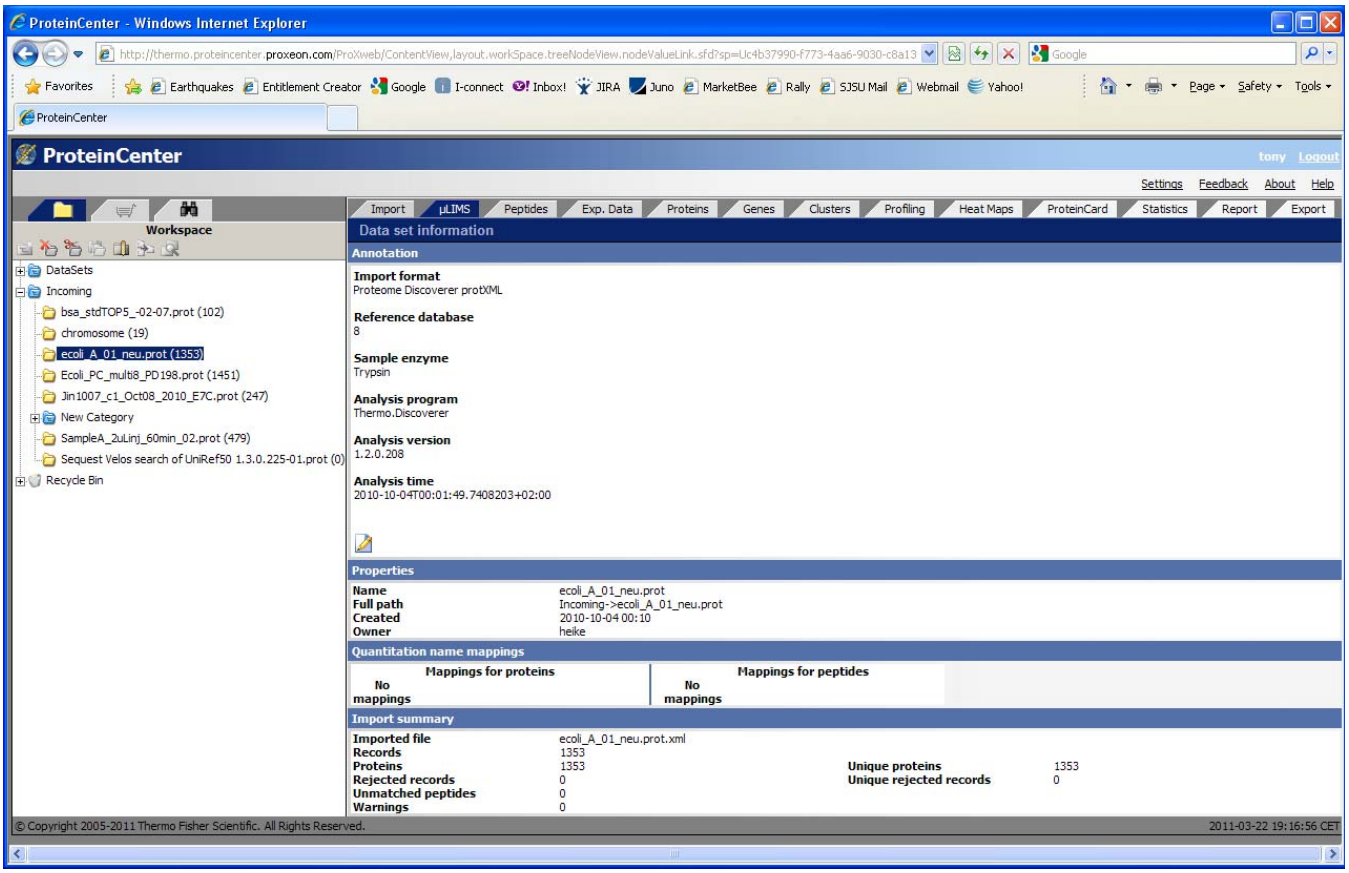
7. If you want to export only the result data from selected protein groups, select the **Checked Protein Groups** check box.

   If you do not select Checked Protein Groups, the Proteome Discoverer application exports the result data of all protein groups.

8. Click **Export**.

   After the Proteome Discoverer application exports the data set to ProteinCenter, you can log in to your ProteinCenter account. The uploaded data set appears under the Incoming node in the ProteinCenter window, as shown in Figure 239.

Figure 239. Uploaded data set in ProteinCenter window



ProteinCenter Page Parameters

Table 69 lists the parameters on the ProteinCenter page of the Options dialog box.

Table 69. ProteinCenter page parameters

Command or Option	Description
Upload URL	Specifies the URL of the ProteinCenter server to use to upload your search results.
User Name	Specifies the user name of your ProteinCenter user account.
Password	Specifies the password of your ProteinCenter user account.
Test	Verifies that the URL that you specified in the URL box is valid. However, it does not verify that the user name and password are valid.

Accessing ProteinCards

You can access the data in ProteinCenter through the ProteinCard for each protein. In ProteinCard, a protein is considered a specific amino acid sequence in a given species.

❖ To access the data in ProteinCenter

1. Double-click a grid cell on the Proteins page of the MSF file, or select a cell and choose **Search Report > Show Protein ID Details**, or click the **Show Protein/Peptide ID Details** icon,  .

   You might experience a short delay as the Proteome Discoverer application accesses the URL.

2. In the Protein Identification Details dialog box, click the **ProteinCard** tab.

   After loading data from the ProteinCenter server, the Proteome Discoverer application displays the data in the ProteinCard tab. By default, it shows the General tab, shown in Figure 241 on page 451.

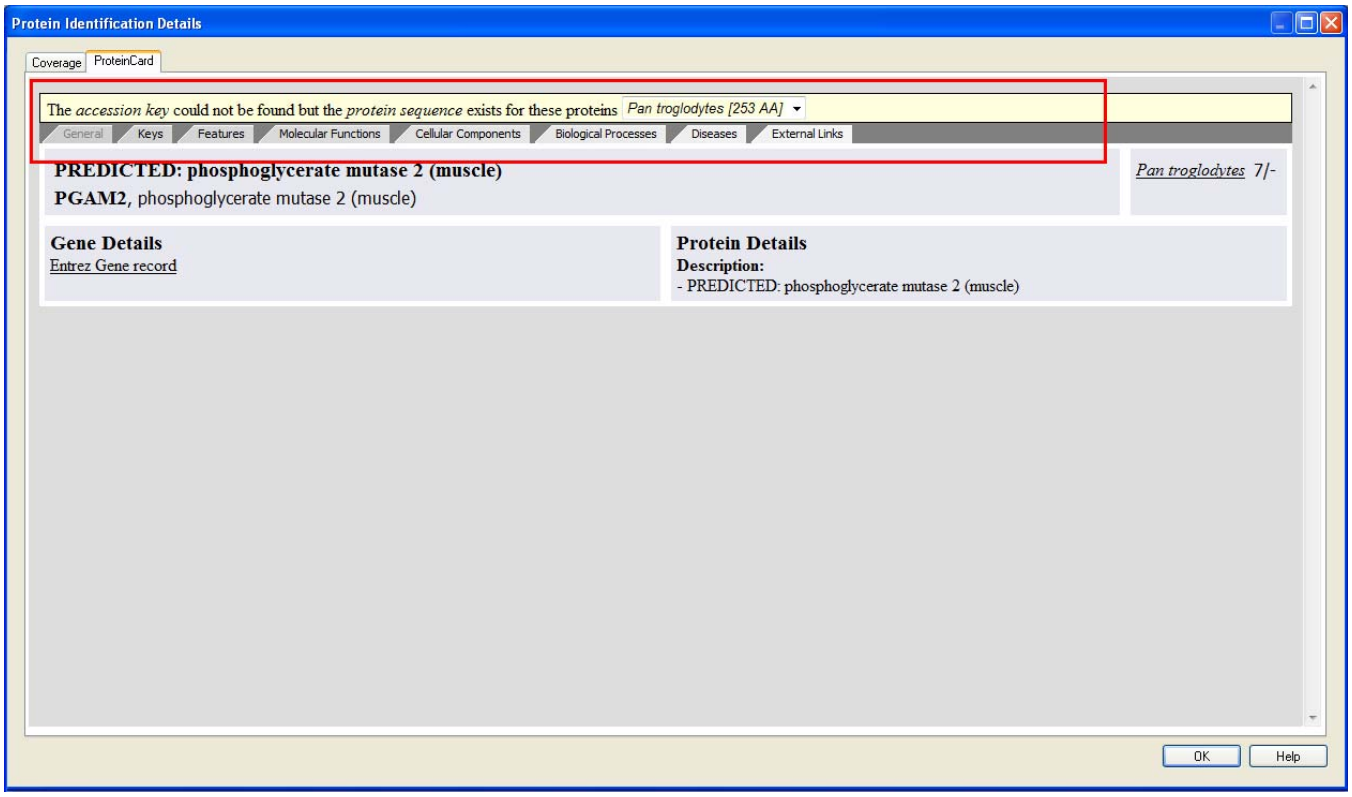
3. Click on the tab of the page containing the information that you are seeking:

   - General Page
   - Keys Page
   - Molecular Functions Page
   - Molecular Functions Page
   - Cellular Components Page
   - Biological Processes Page
   - Diseases Page
   - External Links Page

4. Click **OK** to close the Protein Identification Details dialog box.

If the entire protein is not found in ProteinCenter but a protein with the same sequence exists, the ProteinCard displays a warning that the displayed information is from a protein with different accession, as shown in . If there is more than one protein with the same sequence but from different organisms, an additional list box appears so that you can select the correct species.

Figure 240. Warning displayed for protein with different accession



ProteinCard Parameters

The ProteinCard page of the Protein Identification Details dialog box contains the following pages.

General Page

The General page of the ProteinCard, shown in Figure 241, displays information about the protein: its name, its description, its function, the keywords that produce it in a database search, and the gene that ultimately directs the protein's synthesis through RNA.

Figure 241. General page of the ProteinCard

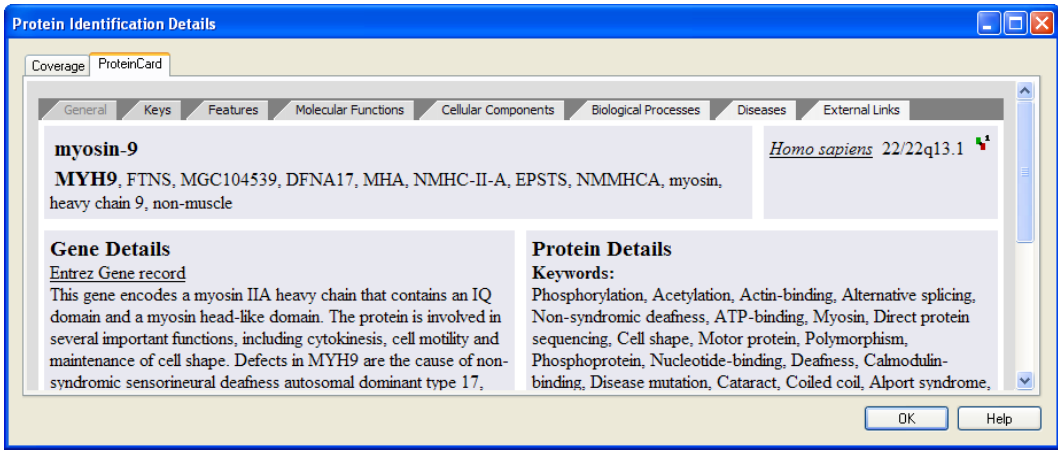


Table 71 lists the parameters on the General page of the ProteinCard page.

Table 70. Parameters on the General page of the ProteinCard page

Command	Description
Top area	Displays the protein name in bold font on the first line. The second line in bold font is the official symbol of the gene that ultimately directs the synthesis of the protein through RNA, and the text following it is the alternative name or names of the gene.
Top right area	Displays the name of the species that contains the gene that ultimately directs the synthesis of this protein through RNA, the number of the chromosome that the gene resides on, and the location of the chromosome that the gene resides on. The name of the species is linked to the National Center for Biotechnology Information (NCBI) taxonomy browser.
Gene Details area	Displays information about the gene that directs the synthesis of the protein. If no information about the gene is available, a link to the Entrez Gene Web site is given.
Protein Details area	Lists the keywords that produce this protein in a database search, the functions of the protein, and a description of the protein.

Keys Page

The Keys page of the ProteinCard, shown in Figure 242, lists all the accession keys for a given protein.

Figure 242. Keys page of the ProteinCard page

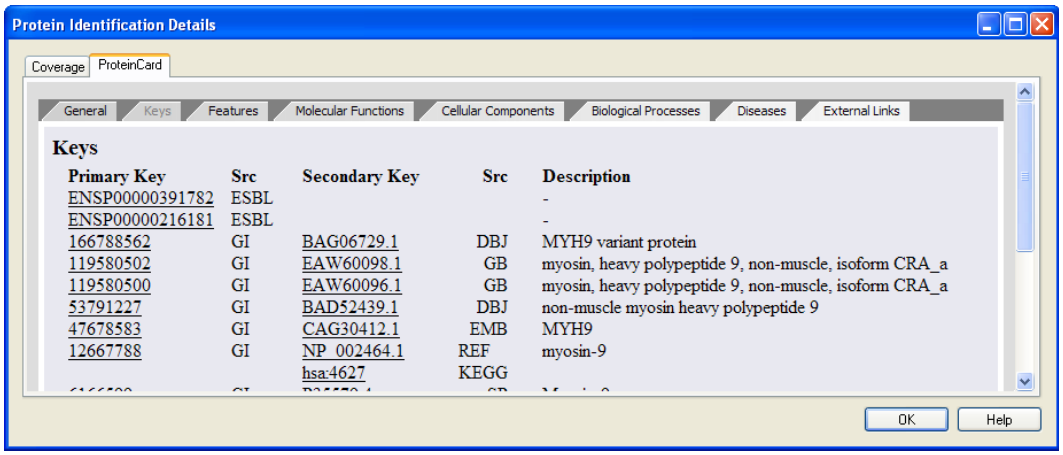


Table 71 lists the parameters on the Keys page of the ProteinCard page.

Table 71. Parameters on the Keys page of the ProteinCard page

Command	Description
Primary Key	Lists the accession key of the database that the sequence was imported from. It is linked to the original database records in the source database, such as Ensembl, SGD, NRDB, IPI or UniProt. The preferred type of accession is emphasized.
Src	Specifies the abbreviation of the primary source database.
Secondary Key	Lists the secondary accession key which is either an alternative key used in the source database or the key of the original database.
Src	Specifies the abbreviation of the secondary source database.
Description	Displays the original description for the original database entry.

An exclamation mark flags outdated protein keys, and the keys are linked to the outdating history in their respective source database.

Features Page

The Features page of the ProteinCard page, shown in Figure 243, includes a selection of sequence features from UniProt, from various conserved domain predictions, and from the computational enrichment undertaken by ProteinCenter. (Computational enrichment refers to information that has no experimental evidence but was found by using a computer prediction program.) The features are sorted according to their start positions in the protein sequence.

Figure 243. Features page of the ProteinCard page

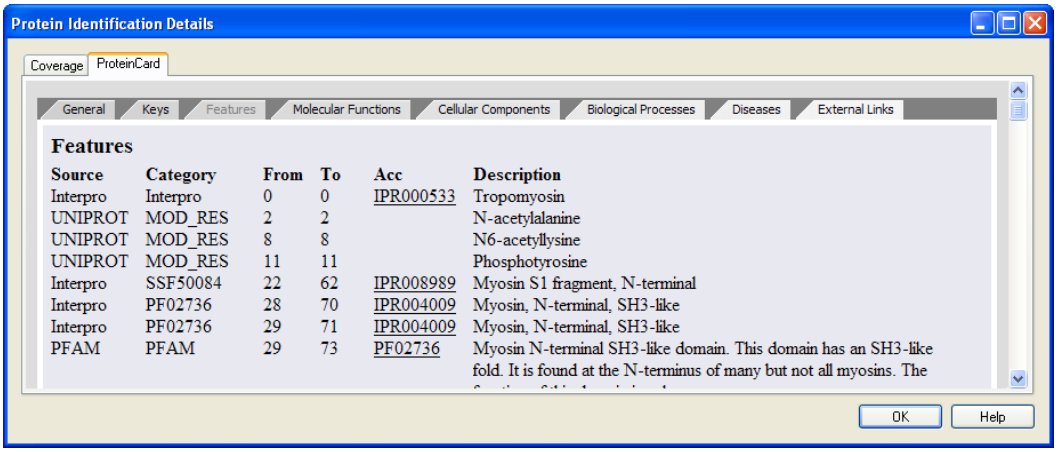


Table 72 lists the parameters on the Features page of the ProteinCard page.

Table 72. Parameters on the Features page of the ProteinCard page (Sheet 1 of 2)

Command	Description
Source	Specifies the name of the database that the information about the feature was taken from: • InterPro • Tmap (computational enrichment) • PrediSi (computational enrichment) • Pfam (computational enrichment) • UniProt
Category	Displays the type of information that UniProt, InterPro, and Tmap include for each row. For example, UniProt might include “CARBOHYD” as one of its types of information, and InterPro might include “SSF57184” as one of its types of information.
From	Specifies the start position of the amino acid.
To	Specifies the end position of the amino acid.

Table 72. Parameters on the Features page of the ProteinCard page (Sheet 2 of 2)

Command	Description
Acc	Specifies the accession identifier for the domain linked to InterPro or Pfam.
Description	Describes the feature.

Molecular Functions Page

The Molecular Functions page of the ProteinCard page, shown in Figure 244, summarizes information about the function of the protein. It consolidates GO data and Enzyme Category (EC) information. The EC designation indicates whether a protein has been categorized with a certain enzyme function.

Figure 244. Molecular Functions page of the ProteinCard page

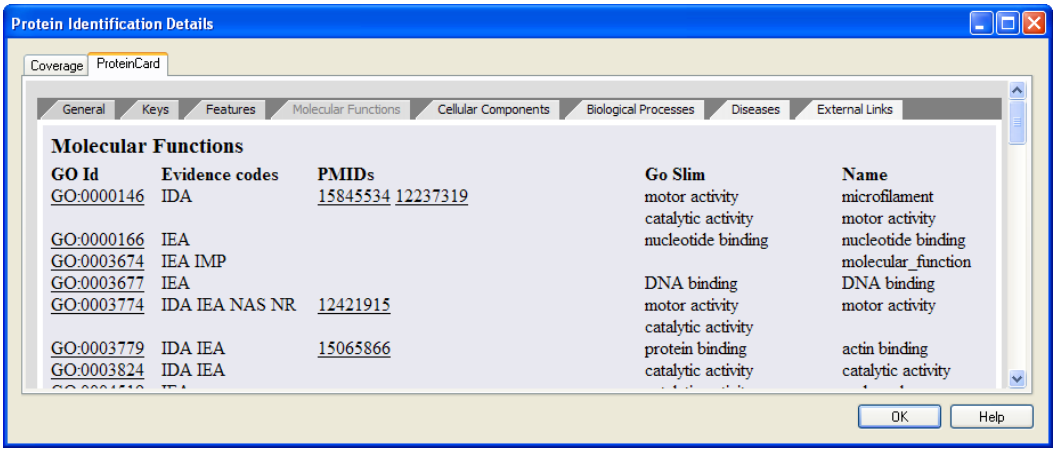


Table 73 lists the parameters on the Molecular Functions page of the ProteinCard page.

Table 73. Parameters on the Molecular Functions page of the ProteinCard page (Sheet 1 of 2)

Command	Description
GO Id	Lists the GO code for each of the protein's molecular functions. Each code is linked to the QuickGO browser of the European Bioinformatics Institute (EBI), which hosts several databases and services.
Evidence Codes	Lists the evidence codes for each of the protein's molecular functions for GO annotation. Evidence codes describe how the GO information was proven—for example, by computer prediction or by experiment.

Table 73. Parameters on the Molecular Functions page of the ProteinCard page (Sheet 2 of 2)

Command	Description
PMIDs	Lists the molecular function codes in the PubMed database, which is maintained by the U.S. National Library of Medicine (NLM) and the National Institutes of Health (NIH). Each code is linked to the PubMed browser.
Go Slim	Specifies the basic GO Slim category for the GO term. GO Slim categories are reduced versions of the GO ontologies containing a subset of the terms in the entire GO database. They give a broad overview of the ontology content without the detail of the specific fine-grained terms. Table 76 provides the Go Slim categories for molecular functions.
Name	Describes the molecular function for a GO term. This description is created by the GO consortium.

Enzymes with an EC number for IUBMB Enzyme Nomenclature are displayed with links to detailed information at the International Union of Biochemistry and Molecular Biology.

Cellular Components Page

The Cellular Components page of the ProteinCard page, shown in Figure 245, summarizes information about where the protein carries out its function in the cell.

Figure 245. Cellular Components page of the ProteinCard page

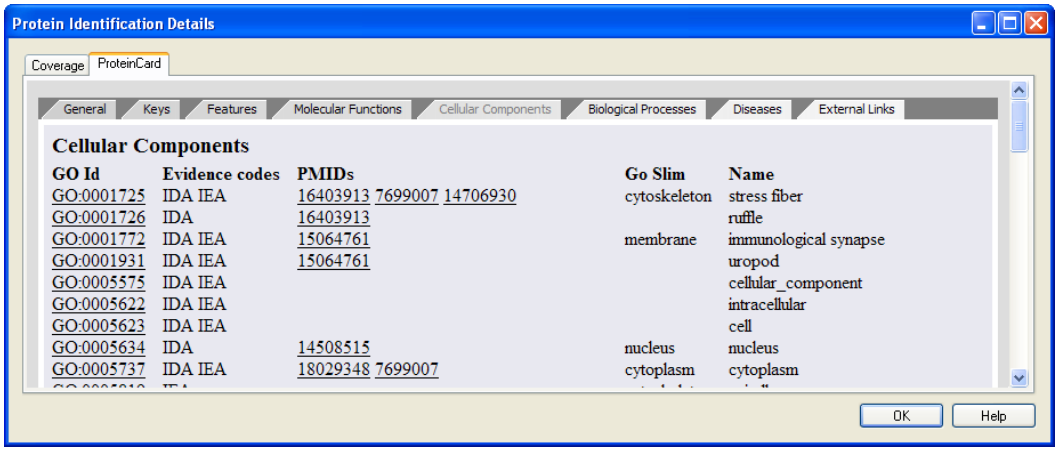


Table 74 lists the parameters on the Cellular Components page of the ProteinCard page.

Table 74. Parameters on the Cellular Components page of the ProteinCard page

Command	Description
GO Id	Lists the GO code for each of the protein's molecular functions. Each code is linked to the QuickGO browser of the EBI, which hosts a number of databases and services.
Evidence Codes	Lists the evidence codes for each of the protein's cellular components for GO annotation. Evidence codes describe how the GO information was proven—for example, by computer prediction or by experiment.
PMIDs	Lists the cellular component codes in the PubMed database, which is maintained by the NLM and the NIH. Each code is linked to the PubMed browser.
Go Slim	Specifies the basic GO Slim category for the GO term. GO Slim categories are reduced versions of the GO ontologies containing a subset of the terms in the entire GO database. They give a broad overview of the ontology content without the detail of the specific fine-grained terms. Table 77 provides the Go Slim categories for cellular components.
Name	Describes the cellular component for a GO term. This description is created by the GO consortium.

Enzymes with an EC number for IUBMB Enzyme Nomenclature are displayed with links to detailed information at the International Union of Biochemistry and Molecular Biology.

Biological Processes Page

The Biological Processes page of the ProteinCard page, shown in Figure 246, summarizes information about the biological processes that the protein is a part of.

Figure 246. Biological Processes page of the ProteinCard page

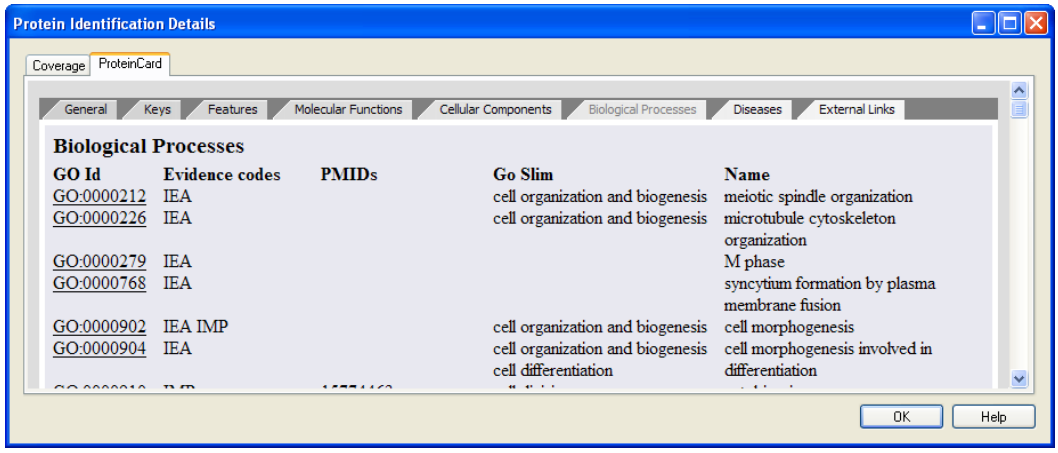


Table 75 lists the parameters on the Biological Processes page of the ProteinCard page.

Table 75. Parameters on the Biological Processes page of the ProteinCard page

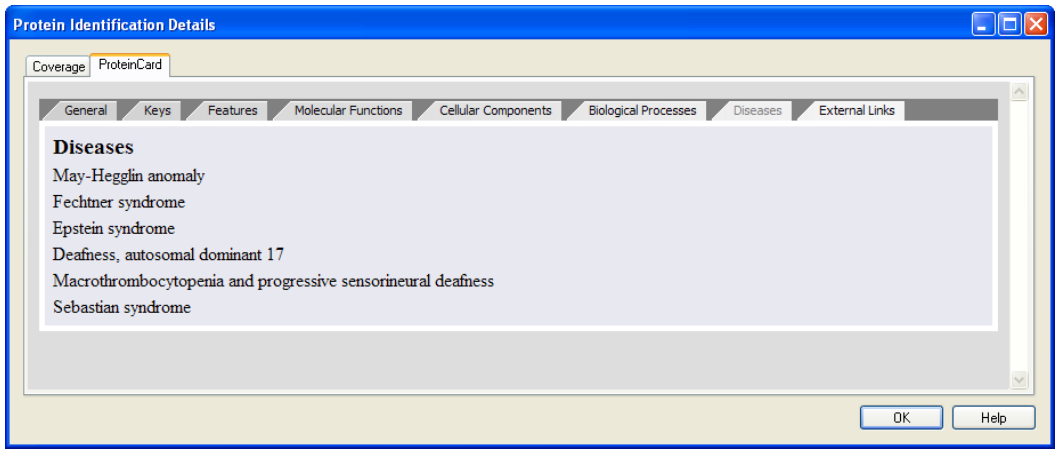
Command	Description
GO Id	Lists the GO code for each of the protein's molecular functions. Each code is linked to the QuickGO browser of the EBI, which hosts a number of databases and services.
Evidence Codes	Lists the evidence codes for each of the protein's biological processes for GO annotation. Evidence codes describe how the GO information was proven—for example, by computer prediction or by experiment.
PMIDs	Lists the biological process codes in the PubMed database, which is maintained by the NLM and the NIH. Each code is linked to the PubMed browser.
Go Slim	Specifies the basic GO Slim category for the GO term. GO Slim categories are reduced versions of the GO ontologies containing a subset of the terms in the entire GO database. They give a broad overview of the ontology content without the detail of the specific fine-grained terms. Table 78 provides the Go Slim categories for biological components.
Name	Describes the biological process for a GO term. This description is created by the GO consortium.

Enzymes with an EC number for IUBMB Enzyme Nomenclature are displayed with links to detailed information at the International Union of Biochemistry and Molecular Biology.

Diseases Page

The Diseases page of the ProteinCard page, shown in Figure 247, lists the diseases that the selected protein is associated with.

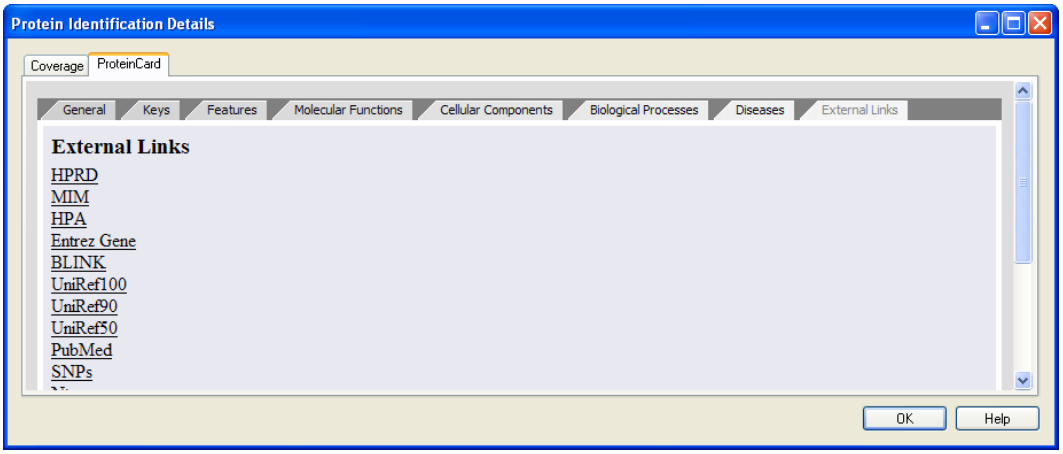
Figure 247. Diseases page of the ProteinCard page



External Links Page

The External Links page of the ProteinCard page, shown in Figure 248, lists the Web links to resources containing information about the protein.

Figure 248. External Links page of the ProteinCard page



Click the appropriate link to open the browser for the database. The external links contains links to resources containing information about the respective protein.

GO Slim Categories

This section defines the GO Slim terms for molecular functions, cellular components, and biological processes.

GO Slim Categories for Molecular Functions

Table 76 describes the GO Slim categories for molecular functions.

Table 76. GO Slim categories for molecular functions (Sheet 1 of 2)

GO Slim molecular function	Description
Antioxidant activity	Inhibition of the reactions brought about by dioxygen (O2) or peroxides. Usually the antioxidant is effective because it can be more easily oxidized than the substance protected. The term is often applied to components that can trap free radicals, breaking the chain reaction that normally leads to extensive biological damage.
Catalytic activity	Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates and are usually composed wholly or largely of protein.
DNA binding	Selective interaction with DNA (deoxyribonucleic acid).
Enzyme regulator activity	Modulation of an enzyme.
Metal ion binding	Selective interaction with any metal ion.
Motor activity	Catalysis of movement along a polymeric molecule such as a microfilament or microtubule, coupled to the hydrolysis of a nucleoside triphosphate.
Nucleotide binding	Selective interaction with a nucleotide, which is any compound consisting of a nucleoside that is esterified with (ortho)phosphate or an oligophosphate at any hydroxyl group on the ribose or deoxyribose moiety.
Protein binding	Selective interaction with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).
Receptor activity	The mediation by protein or gene products of a signal from the extracellular environment to a intracellular messenger.

Table 76. GO Slim categories for molecular functions (Sheet 2 of 2)

GO Slim molecular function	Description
RNA binding	Selective interaction with an RNA molecule or a portion of it.
Signal transducer activity	Mediation of the transfer of a signal from the outside to the inside of a cell by means other than the introduction of the signal molecule itself into the cell.
Structural molecule activity	The action of a molecule that contributes to the structural integrity of a complex or assembly within or outside a cell.
Transcription regulator activity	Activity that plays a role in regulating transcription; it might bind a promoter or enhancer DNA sequence or interact with a DNA-binding transcription factor.
Translation regulator activity	The initiation, activation, perpetuation, repression, or termination of polypeptide synthesis at the ribosome.
Transporter activity	Activity that enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within, or between cells.

GO Slim Categories for Cellular Components

Table 77 describes the GO Slim categories for cellular components.

Table 77. GO Slim categories for cellular components (Sheet 1 of 4)

GO Slim cellular component	Description
Cell surface	Proteins that are attached to the external part of the cell wall, cell membrane, or both.
Chromosome	A structure composed of a very long molecule of DNA and associated proteins (for example, histones) that carry hereditary information.
Cytoplasm	All of the contents of a cell excluding the plasma membrane and nucleus but including other subcellular structures.

Table 77. GO Slim categories for cellular components (Sheet 2 of 4)

GO Slim cellular component	Description
Cytoskeleton	Any of the various filamentous elements that form the internal framework of cells and that typically remain after treatment of the cells with mild detergent to remove membrane constituents and soluble components of the cytoplasm. The term embraces intermediate filaments, microfilaments, microtubules, the microtrabecular lattice, and other structures characterized by a polymeric filamentous nature and long-range order within the cell. The various elements of the cytoskeleton not only serve in the maintenance of cellular shape but also have roles in other cellular functions, including cellular movement, cell division, endocytosis, and movement of organelles.
Cytosol	That part of the cytoplasm that does not contain membranous or particulate subcellular components.
Endosome	A membrane-bound organelle that carries materials newly ingested by endocytosis. It passes many of the materials to lysosomes for degradation.
Endoplasmatic reticulum	The irregular network of unit membranes, visible only by electron microscopy, that occurs in the cytoplasm of many eukaryotic cells. The membranes form a complex meshwork of tubular channels, which are often expanded into slit-like cavities called cisternae. The endoplasmatic reticulum takes two forms, rough (or granular), with ribosomes adhering to the outer surface, and smooth, with no ribosomes attached.
Extracellular	The space external to the outermost structure of a cell. For cells without external protective or external encapsulating structures, this term refers to the space outside of the plasma membrane. It only applies to proteins that are not attached to the cell surface. It covers the host cell environment outside an intracellular parasite.

Table 77. GO Slim categories for cellular components (Sheet 3 of 4)

GO Slim cellular component	Description
Golgi	A compound membranous cytoplasmic organelle of eukaryotic cells consisting of flattened, ribosome-free vesicles arranged in a more or less regular stack. The Golgi apparatus differs from the endoplasmic reticulum in often having slightly thicker membranes, appearing in sections as a characteristic shallow semicircle so that the convex side (cis or entry face) abuts the endoplasmic reticulum, secretory vesicles emerging from the concave side (trans or exit face). In vertebrate cells, there is usually one such organelle, but in invertebrates and plants, where they are known usually as dictyosomes, there may be several scattered in the cytoplasm. The Golgi apparatus processes proteins produced on the ribosomes of the rough endoplasmic reticulum. Such processing includes modification of the core oligosaccharides of glycoproteins and the sorting and packaging of proteins for transport to a variety of cellular locations.
Membrane	Double layer of lipid molecules that encloses all cells, and, in eukaryotic cells, many organelles. The membrane can be a single or double lipid bilayer. It also includes associated proteins. **Note** This term is not restricted to the plasma membrane but applies to all types of membranes present in the cell, that is, nuclear membranes and mitochondrial membranes.
Mitochondrion	A semiautonomous, self-replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration.
Nucleus	A membrane-bounded organelle of eukaryotic cells in which chromosomes are housed and replicated. In most cells, the nucleus contains all of the cell's chromosomes except the organellar chromosomes and is the site of RNA synthesis and processing. In some species or in specialized cell types, RNA metabolism or DNA replication might be absent.
Spliceosome	A ribonucleoprotein complex containing RNA and small nuclear ribonucleoproteins (snRNPs), which is assembled during the splicing of messenger RNA primary transcript to excise an intron.

Table 77. GO Slim categories for cellular components (Sheet 4 of 4)

GO Slim cellular component	Description
Protein complex	Any protein group composed of two or more subunits, which may or may not be identical. Protein complexes might have other associated non-protein prosthetic groups, such as nucleic acids, metal ions, or carbohydrate groups.
Ribosome	An intracellular organelle, about 200 Angstroms in diameter, consisting of RNA and protein. It is the site of protein biosynthesis resulting from translation of messenger RNA (mRNA).
Vacuole	A closed structure found only in eukaryotic cells, completely surrounded by unit membrane and containing liquid material. Cells contain one or several vacuoles that might have different functions from each other. Vacuoles have a diverse array of functions. They can act as a storage organelle for nutrients or waste products, as a degradative compartment, as a cost-effective way of increasing cell size, and as a homeostatic regulator controlling both the turgor pressure and the pH of the cytosol.
Organelle lumen	The volume enclosed by the membranes of a particular organelle, for example, endoplasmic reticulum lumen or the space between the two lipid bilayers of a double membrane surrounding an organelle (for example, nuclear membrane lumen).

GO Slim Categories for Biological Processes

Table 78 describes the GO Slim categories for biological processes.

Table 78. GO Slim categories for biological processes (Sheet 1 of 3)

Go Slim biological process	Description
Cell communication	Any process that mediates interactions between a cell and its surroundings. Cell communication encompasses interactions such as signaling or attachment between one cell and another cell, between a cell and an extracellular matrix, or between a cell and any other aspect of its environment.
Cell death	The specific activation or halting of processes within a cell so that its vital functions markedly cease, rather than simply deteriorating gradually over time, which culminates in cell death.

Table 78. GO Slim categories for biological processes (Sheet 2 of 3)

Go Slim biological process	Description
Cell differentiation	The process in which relatively unspecialized cells—for example, embryonic or regenerative cells—acquire specialized structural features, functional features, or both that characterize the cells, tissues, or organs of the mature organism or some other relatively stable phase of the organism's life history. Differentiation includes the processes involved in commitment of a cell to a specific fate.
Cell division	The processes resulting in the physical partitioning and separation of a cell into daughter cells.
Cell growth	The process by which a cell irreversibly increases in size over time by accretion and biosynthetic production of matter similar to that already present.
Cell homeostasis	The processes involved in the maintenance of an internal equilibrium at the level of the cell.
Cell motility	Any process involved in the controlled movement of a cell.
Cell organization and biogenesis	A process that is carried out at the cellular level and that results in the formation, arrangement of constituent parts, or disassembly of a cellular component. The process includes the plasma membrane and any external encapsulating structures, such as the cell wall and cell envelope.
Cell proliferation	The multiplication or reproduction of cells, resulting in the rapid expansion of a cell population.
Coagulation	The process by which a fluid solution, or part of it, changes into a solid or semisolid mass.
Conjugation	The union or introduction of genetic information from compatible mating types that results in a genetically different individual. Conjugation requires direct cellular contact between the organisms.
Defense response	Reactions triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention and recovery from the infection caused by the attack.

Table 78. GO Slim categories for biological processes (Sheet 3 of 3)

Go Slim biological process	Description
Development	The biological process whose specific outcome is the progression of an organism over time from an initial condition (for example, a zygote or a young adult) to a later condition (for example, a multicellular animal or an aged adult).
Metabolic process	Processes that cause many of the chemical changes in living organisms, including anabolism and catabolism. Metabolic processes typically transform small molecules but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation.
Regulation of biological process	Any process that modulates the frequency, rate, or extent of a biological process. Biological processes are regulated by many means, for example, control of gene expression, protein modification, or interaction with a protein or substrate molecule.
Reproduction	The production by an organism of new individuals that contain some portion of their genetic material inherited from that organism.
Response to stimulus	A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, and so forth) as a result of a stimulus.
Transport	The directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within, or between cells.

Quantification

This chapter describes how to perform precursor-, reporter-, and peak area-based quantification in the Proteome Discoverer application.

Contents

- Activating the Quantification Menu
- Proteins Included in the Quantification
- Performing Precursor Ion Quantification
- Performing Reporter Ion Quantification
- Performing Peak Area Calculation Quantification
- Setting Up the Quantification Method
- Adding a Quantification Method
- Changing a Quantification Method
- Removing a Quantification Method
- Importing a Quantification Method
- Exporting a Quantification Method
- Summarizing the Quantification
- Displaying Quantification Spectra
- Displaying Quantification Channel Values
- Displaying the Quantification Spectrum Chart
- Using Reporter Ion Isotopic Distribution Values to Correct for Impurities
- Excluding Peptides from the Protein Quantification Results
- Excluding Peptides with High Levels of Co-Isolation
- Calculating Peptide Ratios
- Calculating Protein Ratios from Peptide Ratios

Contents - continued

- Calculating Ratio Count and Variability
- Calculating and Displaying Protein Ratios for Multiconsensus Reports
- Identifying Isotope Patterns in Precursor Ion Quantification
- Troubleshooting Quantification

Activating the Quantification Menu

In the Proteome Discoverer application, the Quantification menu becomes available when you open an MSF file generated by a workflow in the Workflow Editor that includes the Reporter Ions Quantifier node, the Precursor Ions Quantifier node, or the Precursor Ions Area Detector node.

❖ **To activate the Quantification menu**

- Choose **File > Open Report** and follow the procedure given in "Opening MSF Files" on page 215 to open an MSF file containing quantification results.

  The commands on the Quantification menu now become available.

  If you do not have an MSF file containing quantification results, see "Performing Precursor Ion Quantification," "Performing Reporter Ion Quantification" on page 476, or "Performing Peak Area Calculation Quantification" on page 483 for instructions on creating one.

Proteins Included in the Quantification

To determine the proteins to include in the quantification, the Proteome Discoverer application first creates protein groups from the identified PSMs. When the search results include quantification data, it then performs quantification on all protein groups.

The application calculates the ratio for each of the defined quantification ratios for the protein group as the median of all PSMs belonging to the protein group that are marked as being usable. Whether the application considers a PSM usable is determined by the settings of the Quantification Method Editor dialog box, including two options on the Protein Quantification page, Use Only Unique Peptides and Consider Proteins Groups for Peptide Uniqueness. The Use Only Unique Peptides option includes in the quantification peptides that do not occur in other proteins. The Proteins Groups for Peptide Uniqueness option defines peptide uniqueness on the basis of protein groups rather than individual proteins.

When it determines peptide uniqueness for classification in the PSM Ambiguity column on the Peptides page, the application only considers the PSMs that it considered when creating the protein groups, if you select the Use Only Unique Peptides option. For example, it does not use for quantification a PSM of low confidence that it did not use to create the protein groups.

Performing Precursor Ion Quantification

In precursor ion quantification, also called isotopically labeled quantification, **p**rotein abundance is determined from the relative MS signal intensities of an isotopically labeled sample and an unlabeled control sample. Stable-isotope labeling by amino acid in cell culture (SILAC™) is a proteomics identification and quantification technique that uses in-vivo metabolic labeling to detect differences in the abundance of proteins in multiple samples. It is a type of isotopically labeled quantification, which uses stable (non-radioactive) heavy isotopes as labels. You can also introduce the stable isotopes by chemical labeling at the protein or peptide level with the isotopomeric tags (for example, dimethyl labeling).

The following default quantification methods are available for precursor ion (isotopically labeled) quantification:

- SILAC 2plex (Arg10, Lys6): Uses arginine 10 and lysine 6.
- SILAC 2plex (Arg10, Lys8): Uses arginine 10 and lysine 8.
- SILAC 2plex (Ile6): Uses isoleucine 6.
- SILAC 3plex (Arg6, Lys4|Arg10, Lys8): Uses arginine 10 and lysine 8 for “heavy” labels and arginine 6 and lysine 4 for “medium” labels.
- SILAC 3plex (Arg6, Lys6|Arg10, Lys8): Uses arginine 10 and lysine 8 for “heavy” labels and arginine 6 and lysine 6 for “medium” labels.
- Dimethylation 3plex: Chemically adds isotopically labeled dimethyl groups to the N-terminus and to the ε-amino group of lysine.
- ^{18}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O.

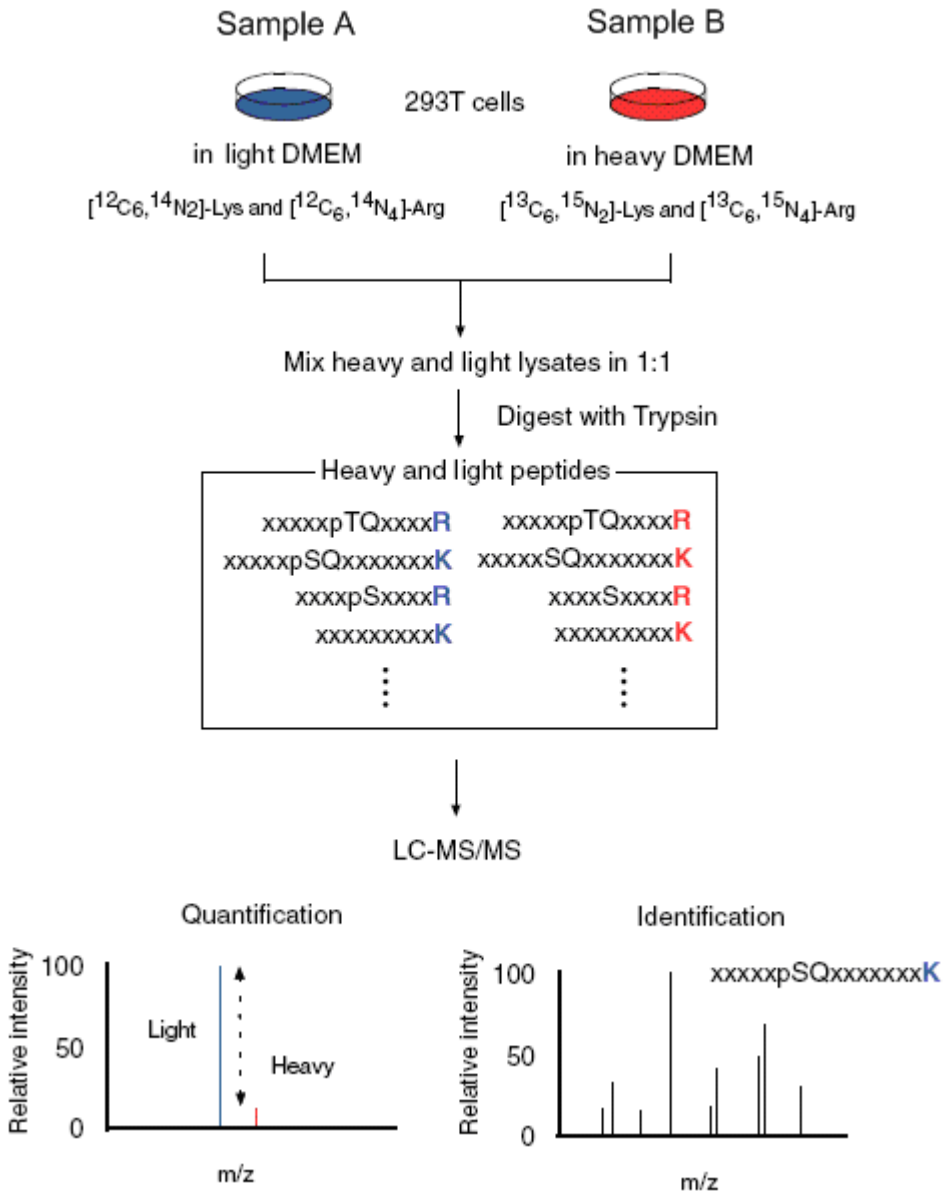
SILAC 2plex Methods

In a typical SILAC quantification experiment, two cell populations grow in media that are deficient in lysine and arginine. One population grows in a medium containing normal (“light”) amino acids, such as lysine ($^{12}C_6{}^{14}N_2$). The other population grows in a medium containing amino acids where stable heavy isotopes, such as lysine 6 ($^{13}C_6{}^{14}N_2$) or lysine 8 ($^{13}C_6{}^{15}N_2$), have been substituted for normal atoms. SILAC quantification usually uses

"heavy" arginine and lysine, because these are the cleavage sites for the generally used trypsin protease. Both populations incorporate these amino acids into proteins through natural cellular protein synthesis. The cells growing in the medium with the heavy isotopes incorporate these isotopes into all of their proteins.

After altering the proteome in one sample through chemical treatment or genetic manipulation, you then combine equal amounts of protein from both cell populations and digest with trypsin before MS analysis. Because peptides labeled with "heavy" and "light" amino acids are chemically identical, they co-elute during reverse-phase chromatographic separation. This means they are detected simultaneously during MS analysis. To determine the average change in protein abundance in the treated sample, you use the relative peak intensities of multiple isotopically distinct peptides from each protein, as shown in Figure 249.

Figure 249. Schematic workflow for SILAC-based peptide and protein quantification



SILAC can differentiate peptides in single MS mode without requiring you to perform tandem mass spectrometry. However, SILAC cannot identify peptides, so you must use tandem mass spectrometry for that purpose.

You can use several SILAC 2plex methods, for example (Arg10, Lys6) and (Arg10, Lys8), to compare two samples.

SILAC 3plex Methods

SILAC 3plex methods are similar to SILAC 2plex methods except, in addition to a "heavy" sample (containing, for example, Arg10 and Lys8), they also use a "medium" sample (containing, for example, Arg6 and Lys4). Protein abundance is determined from the relative MS signal intensities of the heavy sample, medium sample, and a control sample containing "light" (^{12}C and ^{14}N) arginine and lysine.

Dimethylation 3plex Method

The Proteome Discoverer application also includes the dimethylation 3plex method. It is not metabolomic labeling in cell culture but is a form of peptide chemical labeling. This method uses formaldehyde and sodium cyanoborohydride to add dimethyl groups $(CH_3)_2$ to the N-terminus and to the ε-amino group of lysine. By choosing the isotopomers of formaldehyde and sodium cyanoborohydride, you can create light, medium, and heavy labels. For the light label, the (natural-isotope) dimethyl group is $^{12}C_2{}^1H_6$. For the medium label, the dimethyl group is $^{12}C_2{}^2H_4{}^1H_2$, which is 4 Da more massive. For the heavy label, the dimethyl group is $^{13}C_2{}^2H_6$, which is an additional 4 Da more massive.

You can use the dimethylation 3plex method to compare up to three samples. You cannot apply labels to the C terminus, nor to arginine.

^{18}O Labeling Method

The ^{18}O labeling method introduces 2 or 4 Da mass labels through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O.

Creating a Workflow for Precursor Ion Quantification

To use a precursor ion quantification method, you must use a workflow that includes the Precursor Ions Quantifier node.

❖ To create a workflow for precursor ion quantification

Note This procedure uses a SILAC 2plex example.

For instructions on creating a workflow with the Workflow Editor, see "Starting a New Search by Using the Workflow Editor" on page 89.

1. Choose **Workflow Editor > New Workflow**.
2. In the Workflow Editor, drag the **Spectrum Files** node to the workspace.

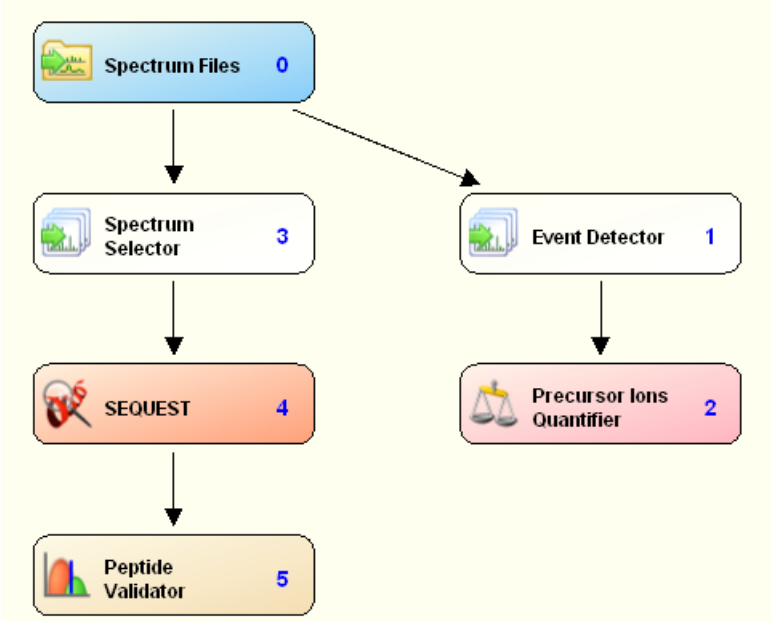
3. If you selected the Spectrum Files node as your input, do the following:
   a. Drag the **Spectrum Selector** node and the **Event Detector** node to the workspace.
   b. Connect the Spectrum Selector node and the Event Detector node to the Spectrum Files node.
4. Drag the **Precursor Ions Quantifier** node to the workspace pane and attach it directly to the Event Detector node.

   The Precursor Ions Quantifier node performs quantification for isotopically labeled amino acids.

   Note You cannot use the Precursor Ions Quantifier node and the Precursor Ions Area Detector node in the same workflow. You cannot use the Reporter Ions Quantifier node in the same workflow with either of these two nodes.
5. Drag the appropriate search engine node (for example, **SEQUEST**) to the workspace pane and attach it to the Spectrum Selector node.
6. Drag the Peptide Validator or the Percolator node to the workspace pane and attach it to the search engine node.

   Figure 250 illustrates the workflow up to this point.

   Figure 250. Beginning of the workflow for precursor ion quantification



7. Add any other nodes that you want and connect them. For general information about creating a workflow in the Workflow Editor, see "Starting a New Search by Using the Workflow Editor" on page 89.
8. In the Parameters pane of the Workflow Editor, click **Show Advanced Parameters**.

9. Click the **Spectrum Files** node and specify the raw file(s) in the Parameters pane.
10. Click the **Event Detector** node and set the parameters for it in the Parameters pane:
    a. In the Mass Precision box, specify the expected standard deviation of the mass precision.

       Three times the standard deviation is used to create extracted ion chromatograms. The minimum value is 1 ppm. The maximum value is 4 ppm. The default is 2 ppm.

    b. In the S/N Threshold box, specify a threshold signal-to-noise value that determines whether the Proteome Discoverer application removes peaks from the spectrum. It removes peaks with a signal-to-noise value below this threshold.

       The minimum value is 1.0, and there is no maximum value. The default is 1.
11. Click the **Spectrum Selector** node, and set the parameters for it in the Parameters pane:
    a. Change the setting in the Max. Precursor Mass box to an appropriate setting. For example, for SILAC 2plex (Arg10, Lys6) quantification, set this option to **6500**.
    b. Change the setting in the S/N Threshold box to an appropriate setting. For example, for SILAC 2plex (Arg10, Lys6) quantification, set this option to **1.5**.

    For other parameters that you can optionally set for the Spectrum Selector node, see "Spectrum Selector Node" on page 640.
12. Click the search engine node (for example, **SEQUEST**), and set the parameters for it in the Parameters pane:
    a. In the Protein Database box, select the **FASTA** database.
    b. In the Dynamic Modifications area, select the dynamic modifications.

       For example, for SILAC 2plex (Arg10, Lys6) quantification, you might select the following two dynamic modifications:

       - **13C(6)/ +6.020 Da (K)**
       - **13C(6)/15N(4)/+10.008 Da (R)**

       If you do not find these labels, you can enable them by following the instructions in "Chemical Modifications" on page 195.
    c. In the Static Modifications area, select the static modifications. For example, for SILAC 2plex (Arg10, Lys6) quantification, select **Carbamidomethyl/ +57.021 Da (C)** in a Static Modification box.
    d. Set any other parameters that you prefer.
13. Set the parameters for all other nodes in the Parameters pane. For information about all the parameters that you can set for each node, see "Workflow Editor Nodes" on page 637. For information on the parameters that you can set for the Precursor Ions Quantifier node, see step 14 of this procedure.

14. Click the **Precursor Ions Quantifier** node and set the parameters for it in the Parameters pane:
    a. Set up the quantification method. Click the **Quantification Method** parameter, and follow the procedure in "Setting Up the Quantification Method" on page 485 to specify the quantification method.
    b. Set the parameters that identify the isotope patterns:
        i. In the RT Tolerance of Isotope Pattern Multiplet [min] box, specify the maximum retention-time tolerance of the A0 peak in the isotope pattern of a quantification multiplet, in minutes. The default is 0.2 minutes.
        ii. In the Single-Peak/Missing Channels Allowed box, specify the maximum number of single-peak or missing quantification channels that are allowed for a valid peptide quantification result. A single-peak quantification channel is a channel that is identified with just a single peak. The maximum number used will not exceed the number of specified channels. The minimum value is 0 (the default). This value indicates that there are at least two peaks in the quantification channel used for quantification.
15. Choose **Workflow Editor > Start Workflow** or click the **Start Workflow** icon,  .

Performing Reporter Ion Quantification

In contrast to the metabolic labeling used by isotopically labeled precursor ion quantification methods such as SILAC, isobarically labeled reporter ion quantification methods use external reagents, or tags, to enzymatically or chemically label proteins and peptides. Reporter ion quantification uses tags that have the same mass. (A reporter ion is a fragment ion with a tag.) The Proteome Discoverer application supports reporter ion quantification for Tandem Mass Tag (TMT) and Isobaric Tag for Relative and Absolute Quantification (iTRAQ) and any user-defined tags. Identification and quantification with both TMT and iTRAQ are performed in the MS/MS scan.

You can quantify all isobarically labeled samples. For iTRAQ, 4plex and 8plex default methods are available. For TMT, 2plex and 6plex default methods are available. You can also add new methods.

TMT Quantification

TMT quantification is a reproducible, highly accurate quantification method that provides both comparative and absolute MS/MS-based quantification of proteins and peptides in biological samples. TMT tagging produces data to calculate the relative abundances of proteins. You can evaluate differential protein expression in one to six samples in a single experiment.

Each sample is labeled with chemically identical tags before mixing the samples, and a single MS run generates a single peak for each peptide, irrespective of which tag it has been given. Between the normalizer and reporter is a cleavable linker, which breaks during MS/MS. The mass reporter ion is split off and measured by the mass spectrometer.

Only MS/MS fragmentation can differentiate the tagged proteins. The reporter ion, measured by the mass spectrometer, generates a different peak. As a result, the peak height/peak integral for each reporter denotes the relative amount of protein originating from each of the labeled samples.

With the quantification functions in the application, you can set filters to see only unique peptides so that every protein associated with the same peptide is not counted, producing a best-results list of peptides. Filtering the number of proteins can give you a more robust final analysis of your experimental set.

Quantification with TMT tags is no different from quantification with iTRAQ (described in "iTRAQ Quantification" on page 477), except that it uses the following default mass tags by Proteome Sciences PLC:

- TMT 2plex
- TMT 6plex

You can use these default methods to create your own quantification templates. For information on adding quantification methods, see "Changing a Quantification Method" on page 520.

iTRAQ Quantification

iTRAQ is a protein quantification technique that uses isobaric amine-specific, stable isotope reagents to label all peptides in up to eight different samples simultaneously. The labeled peptides from each sample are combined, separated by two-dimensional liquid chromatography, and analyzed with tandem mass spectrometry (MS/MS). The same peptide from each sample appears as a single peak in the MS spectrum. In single MS mode, the differentially labeled versions of a peptide are indistinguishable. In tandem MS mode, which isolates and fragments peptides, each tag generates a unique reporter ion. Protein quantification compares the peak intensity of the reporter ions in the MS/MS spectra to assess the relative abundance of the peptides and therefore the proteins that they are derived from.

iTRAQ includes two default mass tags available from Applied Biosystems (ABI) that you can use to label all peptides:

- iTRAQ 4plex, which is standard
- iTRAQ 8plex

The Proteome Discoverer application includes default quantification methods for processing data from iTRAQ 4plex- and iTRAQ 8plex-labeled samples. You can use these methods to create your own workflow templates. For information on adding quantification methods, see "Changing a Quantification Method" on page 520.

iTRAQ quantification works exactly the same as TMT quantification, except that TMT quantification offers 2plex and 6plex quantification methods, and iTRAQ offers 4plex and 8plex quantification methods.

Creating a Workflow for Reporter Ion Quantification

To use an isobarically labeled reporter ion quantification method, you must open an MSF file generated from a workflow that includes the Reporter Ions Quantifier node.

Setting up the workflow for TMT and iTRAQ quantification is basically the same.

❖ To create a workflow for reporter ion quantification

For instructions on creating a workflow with the Workflow Editor, see "Starting a New Search by Using the Workflow Editor" on page 89.

1. Choose **Workflow Editor > New Workflow**.
2. In the Workflow Editor, drag the **Spectrum Files** or the **SRF File Selector** node to the workspace.
3. If you selected the Spectrum Files node as your input, drag the **Spectrum Selector** node to the workspace, and attach the Spectrum Files node to the Spectrum Selector node.

4. Drag the **Reporter Ions Quantifier** node to the workspace pane, and attach the Spectrum Files node to the Reporter Ions Quantifier node.

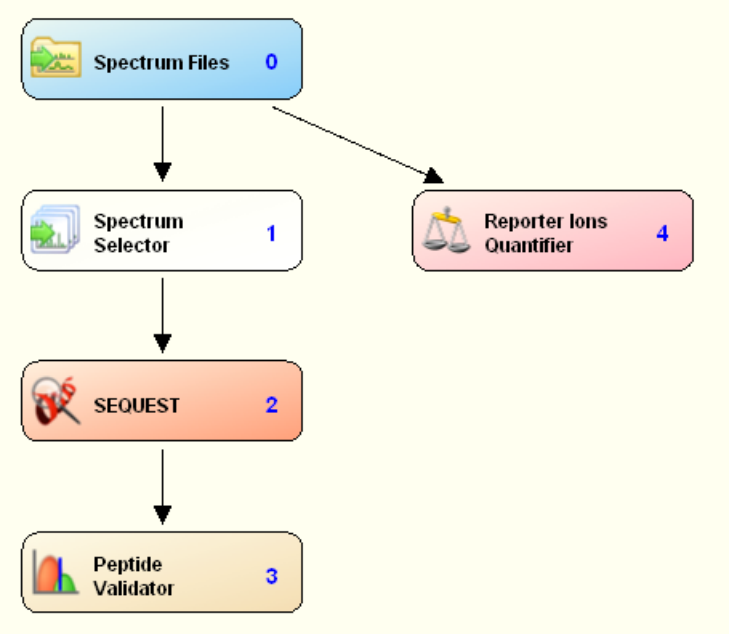
   Note You cannot use the Reporter Ions Quantifier node in the same workflow with either the Precursor Ions Quantifier node or the Precursor Ions Area Detector node.

5. Drag the search engine node that you want (for example, **SEQUEST**) to the workspace pane, and attach the Spectrum Selector node to the search engine node.

6. Drag the Peptide Validator or the Percolator node to the workspace pane and attach it to the search engine node.

   Figure 251 illustrates the workflow up to this point.

   Figure 251. Beginning of the workflow for reporter ion quantification



7. Add any other nodes that you would like and connect them.

   For general information about creating a workflow in the Workflow Editor, see "Starting a New Search by Using the Workflow Editor" on page 89.

8. Click the **Spectrum Files** node and specify the raw file in the Parameters pane.

9. Click the **Spectrum Selector** node, and set the parameters for it in the Parameters pane:

   a. Change the setting in the Total Intensity Threshold box to an appropriate setting.

      For example, for TMT 6plex quantification, you could set this option to **20 000**.

   b. Change the setting in the Minimum Peak Count box to an appropriate setting.

      For example, for TMT 6plex quantification, you could set this option to **200**.

   For other parameters that you can optionally set for the Spectrum Selector node, see "Spectrum Selector Node" on page 640.

10. Click the search engine node (for example, **SEQUEST**), and set the parameters for it in the Parameters pane:
    a. In the Protein Database box, select the **FASTA** database.
    b. In the Dynamic Modifications area, select the dynamic modifications.

       For example, for TMT 6plex quantification, you would select a dynamic modification of **TMT6plex / +229.163 Da (K)**. If you do not find this label, you can enable it by following the instructions in "Chemical Modifications" on page 195.
    c. In the Static Modifications area, select the static modifications. For example, for TMT 6plex quantification, you would select **TMT6plex / +229.163 Da (K)** in the Peptide N-Terminus box.
    d. Set any other parameters that you prefer.
11. Set the parameters for all other nodes in the Parameters pane.

    For information about all the parameters that you can set for each node, see "Workflow Editor Nodes" on page 637. For information on the parameters that you can set for the Reporter Ions Quantifier node, see step 12 of this procedure.
12. Click the **Reporter Ions Quantifier** node and set the parameters for it in the Parameters pane:
    a. Set up the quantification method. Click the **Quantification Method** parameter, and follow the procedure in "Setting Up the Quantification Method" on page 485 to specify the quantification method.
    b. Set the parameters that specify the peak integration:
       i. In the Integration Window Tolerance box, specify the mass-to-charge (*m/z*) window that enables you to look for the reporter peaks. The default is 20 ppm.

ii. In the Integration Method box, select which peak to choose when more than one peak is found inside the integration window:

- (Default) Most Confident Centroid: Lays a Gaussian curve around the target peak (the tag mass) with a sigma value equal to the mass accuracy or integration window. Then the Gaussian curve normalizes all peaks in the window, and the largest is considered to be the most confident peak. This method is also used by the Spectrum Selector node in the Workflow Editor to pick the monoisotopic peak from the survey scan. The only difference is that the Spectrum Selector uses a 3-sigma interval, but Most Confident Centroid uses only a 1-sigma interval. This means the Most Confident Centroid is almost always the largest peak inside the integration window because of the small inclination of the Gaussian curve in the 1-sigma interval.
- Most Intense Centroid: Selects the highest peak.
- Centroid With Smallest Delta Mass: Selects the peak with the smallest deviation from the theoretical mass.
- Centroid Sum: Sums the intensity of all the peaks in the window.

c. Specify the scan event filters:

i. In the Mass Analyzer box, select the type of mass spectrometer used in the acquisition of the spectrum:

- Ion Trap (ITMS)
- (Default) Fourier Transform (FTMS)
- Time of Flight (TOFMS)
- Single Quad (SQMS)
- Triple Quad (TQMS)
- Sector Field (SectorMS)

ii. In the MS Order box, specify the level of tandem mass spectrum to be processed, for example, **MS2** or **MS3**. The default is MS2.

iii. In the Activation Type list, specify the fragmentation method used to activate the scan.

Note You cannot perform TMT quantification on both PQD and HCD scans. You can choose only one activation type.

- CID (Collision-Induced Dissociation)
- ECD (Electron Capture Dissociation)
- ETD (Electron Transfer Dissociation)
- (Default) HCD (High-Energy Collision Dissociation)
- MPD (Multi-Photon Dissociation)
- PQD (Pulsed Q Collision-Induced Dissociation)

For a description of these fragmentation types, see "Fragmentation Methods" on page 8.

13. Choose **Workflow Editor > Start Workflow**, or click the **Start Workflow** icon,  .

Performing TMT Quantification on HCD and CID Scans

If a raw file contains both CID scans for identification and HCD scans for quantification, you can use the following workflow to both quantify the HCD scans and identify peptides in the CID scans, the HCD scans, or both.

❖ To perform TMT Quantification on HCD and CID scans

1. Drag the **Reporter Ions Quantifier** node to the workspace pane and connect it to the workflow.
2. Set the Activation Type parameter for the Reporter Ions Quantifier node to **HCD**.
3. Set the Activation Type parameter for the Spectrum Selector node to **Any**, **Is CID**, **HCD**, or **Is CID**, depending on your method setup and identification strategy.
4. Set all other parameters—modifications, tolerances, FASTA files, and so forth—and choose **Workflow Editor > Start Workflow**, or click the **Start Workflow** icon,  .

Demonstrating How to Create a Workflow for Reporter Ion Quantification

The following demonstration shows you how to set up a workflow for reporter ion quantification and how to specify the quantification method.

Click here to view the demonstration. To see a full-screen version, right-click on the demonstration after you start it and choose **Full Screen Multimedia**.



Performing Peak Area Calculation Quantification

If you want to determine the area for any quantified peptide, you can use peak area calculation quantification. You might want to use this quantification method to obtain an idea of the relative quantities of all peptides in a sample.

If the Proteome Discoverer application calculates peptide areas during processing, it uses them to automatically calculate protein areas for the proteins in the MSF report. It calculates the area of any given protein as the average of the three most abundant distinct peptides identified for the protein.[1] The peptides must have different sequences to be considered distinct. Peptides with different charge states or modification variants of the same sequence are considered the same peptide. If you apply result filters, the application recalculates the protein areas.

❖ To create a workflow for peak area calculation quantification

1. In the Workflow Editor, set up a quantification workflow.

   For instructions on creating a workflow with the Workflow Editor, see "Starting a New Search by Using the Workflow Editor" on page 89.

2. Choose **Workflow Editor > New Workflow**.
3. In the Workflow Editor, drag the **Spectrum Files** node to the workspace.
4. Drag the **Spectrum Selector** node and the **Event Detector** node to the workspace.
5. Connect the Spectrum Selector node and the Event Detector node to the Spectrum Files node.
6. Drag the **Precursor Ions Area Detector** node to the workspace pane and attach it directly to the Event Detector node.

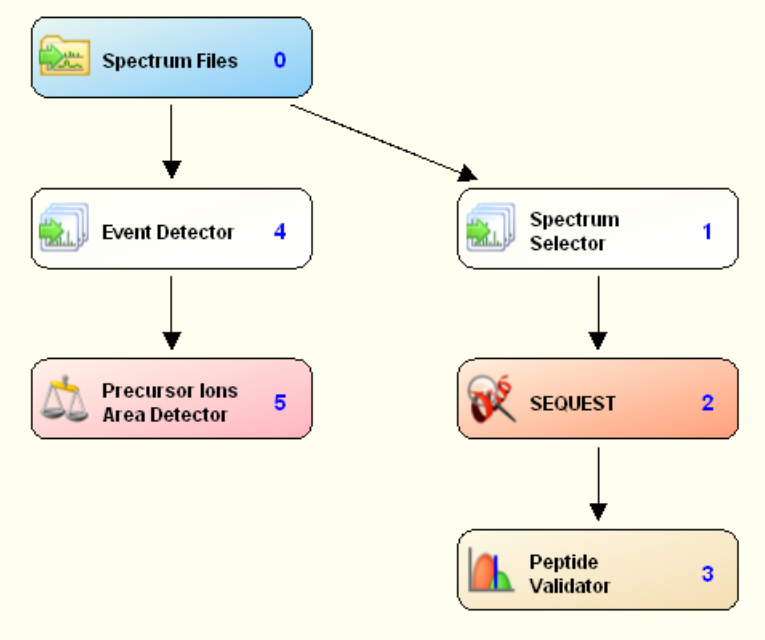
   Note You cannot use the Precursor Ions Area Detector node in the same workflow with the Precursor Ions Quantifier node or the Reporter Ions Quantifier node.

7. Drag the search engine node that you prefer (for example, **SEQUEST**) to the workspace pane and attach it to the Spectrum Selector node.
8. Drag the Peptide Validator or the Percolator node to the workspace pane and attach it to the search engine node.

   Figure 252 illustrates the workflow up to this point.

[1] Silva, J.C.; Gorenstein, M.V.; Li, G.-Z.; Vissers, J.P. C.; and Geromanos, S.J. Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel MS Acquisition. *Molecular & Cellular Proteomics*, **2006**, *5*, 144-156

Figure 252. Beginning of the workflow for area calculation quantification



9. Add any other nodes that you would like and connect them.

   For general information about creating a workflow in the Workflow Editor, see "Starting a New Search by Using the Workflow Editor" on page 89.

10. Click the **Spectrum Files** node and specify the raw file in the Parameters pane.

11. Click the **Event Detector** node and set the parameters for it in the Parameters pane:

    a. In the Mass Precision box, specify the expected standard deviation of the mass precision.

       To create extracted ion chromatograms, use three times the standard deviation. The minimum value is 1 ppm. The maximum value is 4 ppm. The default is 2 ppm.

    b. In the S/N Threshold box, specify a threshold signal-to-noise value that determines whether the Proteome Discoverer application removes peaks from the spectrum.

       The application removes peaks with a signal-to-noise value below this threshold. The minimum value is 0.0, and there is no maximum value. The default is 1.

12. Click the **Spectrum Selector** node, and set the parameters for it in the Parameters pane:

    a. Change the setting in the Max. Precursor Mass box to an appropriate setting. For example, you could set this option to **6500**.

    b. Change the setting in the S/N Threshold box to an appropriate setting. For example, you could set this option to **1.5**.

    For other parameters that you can optionally set for the Spectrum Selector node, see "Spectrum Selector Node" on page 640.

13. Click the search engine node (for example, **SEQUEST**), and set the parameters for it in the Parameters pane:
    a. In the Protein Database box, select an appropriate **FASTA** database.
    b. In the Dynamic Modifications area, select the dynamic modifications.

       For example, you might select the **Oxidation+15.995 Da (M)** dynamic modification. If you do not find this label, you can enable it by following the instructions in "Chemical Modifications" on page 195.
    c. In the Static Modifications area, select the static modifications.

       For example, you might select **Carbamidomethyl / +57.021 Da (C)** in the Static Modification box.
    d. Set any other parameters as needed.
14. Set the parameters for all other nodes in the Parameters pane.

    For information about all the parameters that you can set for each node, see "Workflow Editor Nodes" on page 637. For information on the parameters that you can set for the Precursor Ions Quantifier node, see step 14 of this procedure.
15. Choose **Workflow Editor > Start Workflow**, or click the **Start Workflow** icon,  .

Setting Up the Quantification Method

Setting up the quantification method is similar for both precursor ion quantification and reporter ion quantification. Both methods use values called quantification (quan) channels as the basis for the ratio reporting. You do not need to set up a quantification method for peak area calculation quantification.

For reporter ion quantification, a quantification channel is one of several masses, states, or tags (depending on which quantification method you use) for which you measure a quantification value. The Proteome Discoverer application calculates the reported quantification ratios from the quantification values of the different quantification channels. For example, for iTRAQ 4plex, the different reporter tags (114, 115, 116, 117) are the four quantification channels of the iTRAQ 4plex method. The application calculates the ratios from the detected quantification values of the four quantification channels.

For precursor ion quantification, a quantification channel is one of the different possible labeling states of a peptide corresponding to the different heavy amino acids used in the cell cultures. For example, the SILAC 2plex methods are normally used with two quantification channels named "light" and "heavy." The light quantification channel uses the natural isotopes of lysine ($^{12}C_6{}^{14}N_2$) and arginine ($^{12}C_6{}^{14}N_4$). In the heavy quantification channel, arginine 10 ($^{13}C_6{}^{15}N_4$) replaces all arginines, and either lysine 6 ($^{13}C_6{}^{14}N_2$) or lysine 8 ($^{13}C_6{}^{15}N_2$) replaces all lysines.

❖ To set up the quantification method

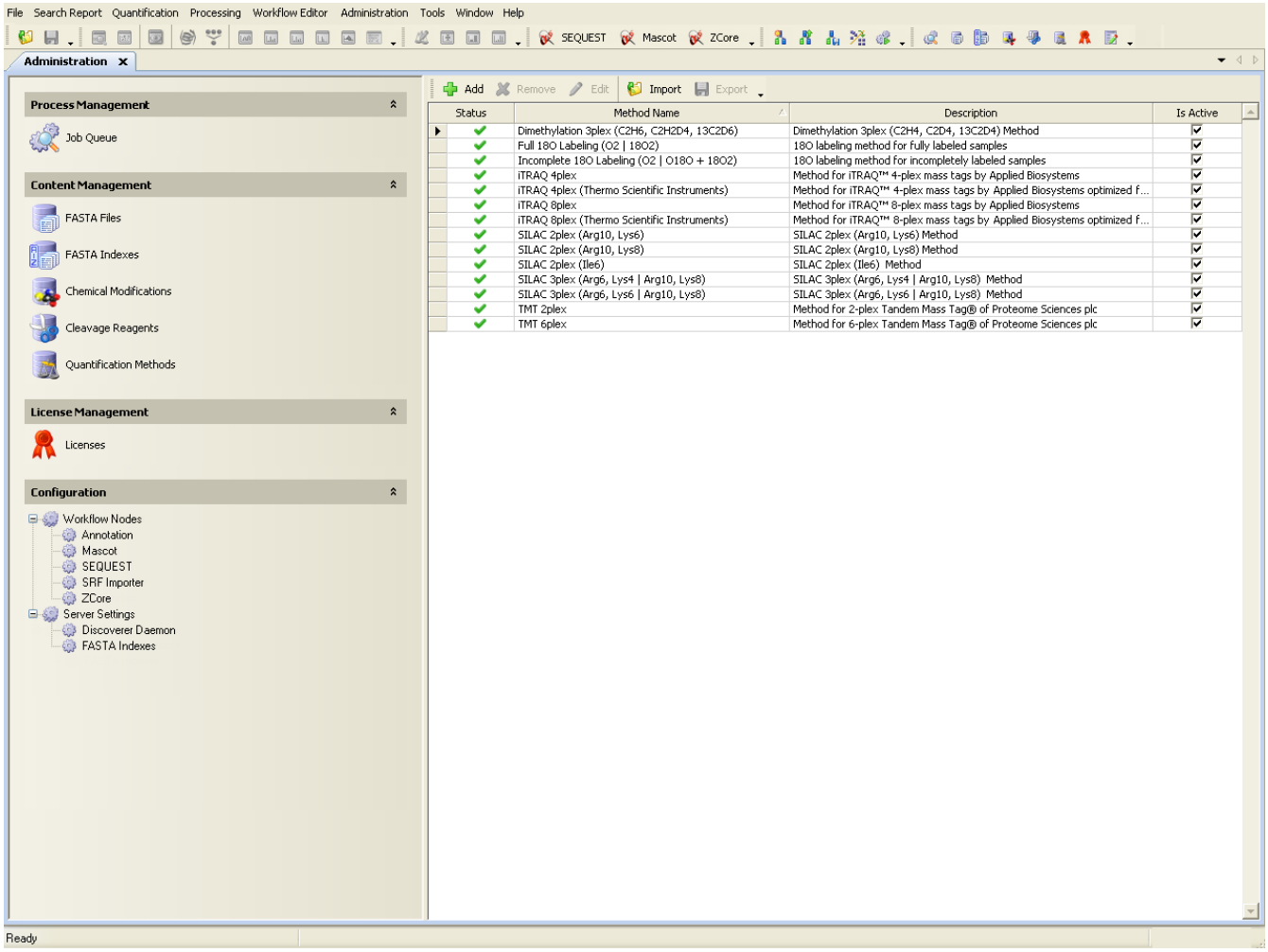
1. Set up a search by following the instructions in "Starting a New Search by Using the Workflow Editor" on page 89.
2. In the workspace pane of the Workflow Editor, add the **Precursor Ions Quantifier** node for precursor ion quantification or the **Reporter Ions Quantifier** node for reporter ion quantification.
3. Click the **Precursor Ions Quantifier** node or the **Reporter Ions Quantifier** node, and in the Quantification Method box, click the browse button (**...**) that appears.

   The Quantification Method Editor dialog box opens to the Quan Channels page. Use this dialog box to set up the quantification method.

   You can also access the Quantification Method Editor dialog box by choosing **Administration > Maintain Quantification Methods** or by clicking the **Maintain Quantification Methods** icon,  , to open the Quantification Methods view, shown in Figure 253. This view lists all of the available methods for both precursor ion and reporter ion quantification.

   Double-click the appropriate method in the Method Name column.

Figure 253. Quantification Methods view



4. From the list at the top of the Quantification Method Editor dialog box, select the quantification method.

   For precursor ion quantification, you can choose from the following methods when you initially set up a workflow and first access the Quantification Method Editor dialog box:

   - SILAC 2plex (Arg10, Lys6): Uses arginine 10 and lysine 6.
   - SILAC 2plex (Arg10, Lys8): Uses arginine 10 and lysine 8.
   - SILAC 2plex (Ile6): Uses isoleucine 6.
   - SILAC 3plex (Arg6, Lys4|Arg10, Lys8): Uses arginine 10 and lysine 8 for “heavy” labels and arginine 6 and lysine 4 for “medium” labels.
   - SILAC 3plex (Arg6, Lys6|Arg10, Lys8): Uses arginine 10 and lysine 8 for “heavy” labels and arginine 6 and lysine 6 for “medium” labels.

- Dimethylation 3plex: Chemically adds isotopically labeled dimethyl groups to the N-terminus and to the ε-amino group of lysine.
- ^{18}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O.

For more information on these methods, see "Performing Precursor Ion Quantification" on page 469.

For reporter ion quantification, you can choose from the following methods when you initially set up a workflow and first access the Quantification Method Editor dialog box:

- iTRAQ 4plex
- iTRAQ 4plex (Thermo Scientific Instruments)
- iTRAQ 8plex
- iTRAQ 8plex (Thermo Scientific Instruments)
- TMT 2plex
- TMT 6plex

The two methods labeled "Thermo Scientific Instruments" have purity corrections optimized for the way Thermo Scientific mass spectrometers process samples and produce data.

For more information on these methods, see "Performing Reporter Ion Quantification" on page 476.

Specifying the Quantification Channels

The first step in setting up the quantification is to specify the quantification channels to use. This process includes a validation step. For precursor ion quantification, the validation step ensures that each peptide is in a valid labeling state according to the labels for the different channels, as defined in the quantification method. For reporter ion quantification, the validation step ensures that only peptides that have one of the specified reporter labels as a modification are considered for protein quantification.

The process of specifying label modifications is similar for precursor ion quantification and reporter ion quantification, but it also has some differences:

- For precursor ion quantification, you specify the label modifications for each quantification channel. For reporter ion quantification, you set the label modifications for the whole method.
- For precursor ion quantification, specifying the label modifications for quantification channels other than the unlabeled channel is mandatory. For reporter ion quantification, specifying the label modifications is optional because the information about the modification of the peptides is not necessary for processing the data. It is only used to verify the peptides when the Proteome Discoverer application loads the reports.

When you specify at least one of the label modifications in the quantification method, the Proteome Discoverer application verifies that each identified peptide has at least one of the specified modifications. It does not matter if only one terminal or only one residue is modified with the specified label modification.

- When the application identifies a peptide with none of the specified label modifications, this peptide cannot be the source of reporter peaks in the MS/MS spectra. As a result, the application marks the peptide "No Quan Labels" in the MSF report. It does not use these peptides when it calculates the protein quantification values from the peptides.
- When the application finds a peptide that does not have an iTRAQ or TMT label as a modification, even though reporter ions were present, it leaves the *Ratio* columns blank.

When you install the Proteome Discoverer application, the default methods for TMT and iTRAQ include the correct label modification. The application does not automatically update already existing reporter methods; you must manually specify the label modifications.

When you open old MSF files that contain reporter quantification data, the label modifications of the quantification method of the MSF file appear as None on the Quan Channels page of the Quantification Method Editor dialog box. You can manually specify the label modification, which then triggers the validation of the peptides, and save the change in the quantification method in the MSF file.

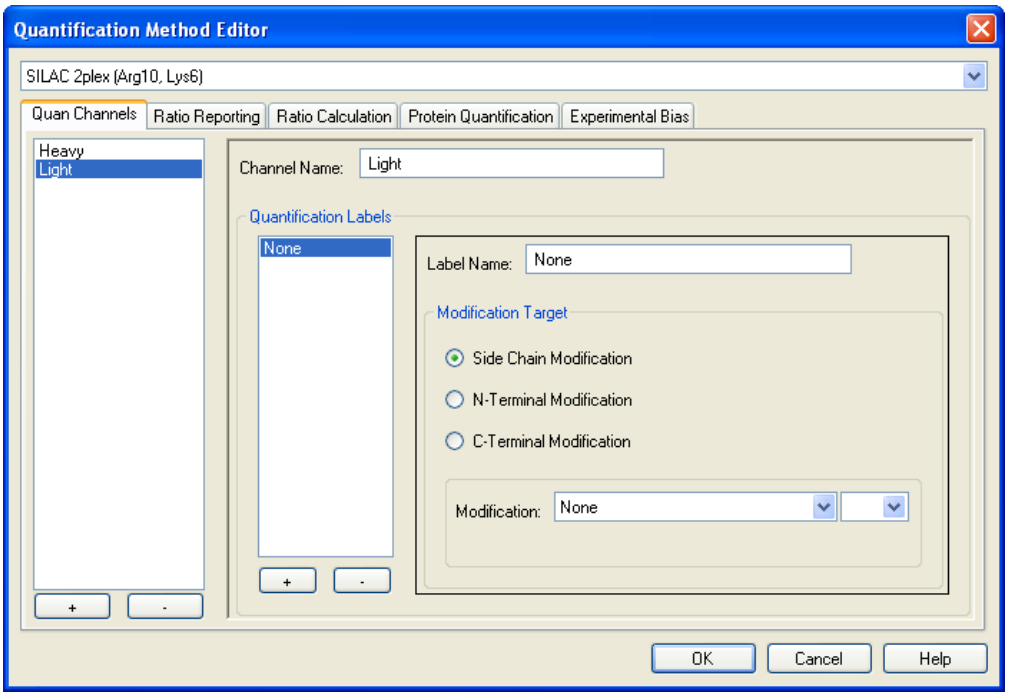
When you do not set the label modifications on the Quan Channels page, the Proteome Discoverer application does not perform the validation.

The process of specifying quantification channels for precursor ion quantification is slightly different from the process of specifying label modifications for reporter ion quantification.

❖ To specify quantification channels for precursor ion quantification

1. Click the **Quan Channels** tab of the Quantification Method Editor dialog box, shown in Figure 254, if it is not already selected.

Figure 254. Quan Channels page of the Quantification Method Editor dialog box for precursor ion quantification



2. In the top list, select the name of the labeling method to use.

   When you create a new workflow and first access the Quantification Method Editor dialog box to set up a quantification method, the default methods available in the top list of the Quan Channels page are as follows:

   - SILAC 2plex (Arg10, Lys6)
   - SILAC 2plex (Arg10, Lys8)
   - SILAC 2plex (Ile6)
   - SILAC 3plex (Arg6, Lys4|Arg10, Lys8)
   - SILAC 3plex (Arg6, Lys6|Arg10, Lys8)
   - Dimethylation 3plex (C2H6, C2D4H2, 13C2D6)
   - ^{18}O labeling

   For a description of these methods, see "Performing Precursor Ion Quantification" on page 469.

   However, after you have chosen a method or set up your own method, only that method appears in the top list of the Quan Channels page after you execute the workflow.

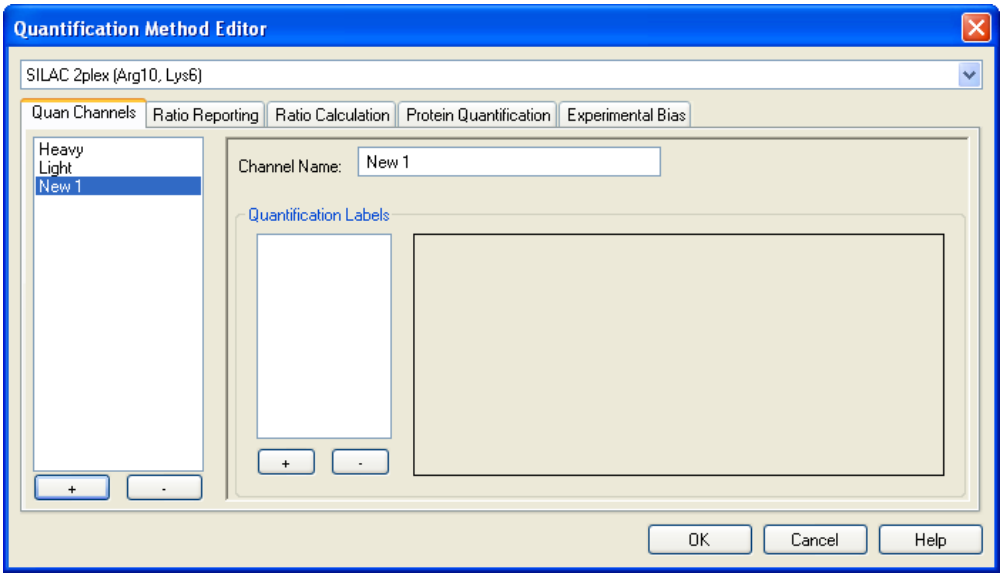
The left box of the Quan Channels page displays two types of default isotopic labels:

- Heavy: Refers to amino acid labels that use heavy isotopes, for example Arg10 and Lys8.
- Medium (3plex methods only): Refers to amino acid labels that use less massive isotopes, for example Arg6 and Lys4.
- Light: Refers to amino acid labels that use normal isotopes.

3. To add a quantification channel, click **+** beneath the list of quantification channels in the box on the left.

   The default name of New *number* now appears in the list of quantification channels and in the Channel Name box, as shown in Figure 255.

Figure 255. New quantification channel on the Quan Channels page



To remove a quantification channel, select the quantification channel in the list of quantification channels and click **–** beneath the list.

4. To specify a name for the new quantification channel, backspace over the default name in the Channel Name box and type the new name. The example in Figure 256 uses Medium.

   The new name now appears in the quantification channel (left) box.

5. To specify a quantification label to assign to a quantification channel, click **+** beneath the Quantification Labels box.

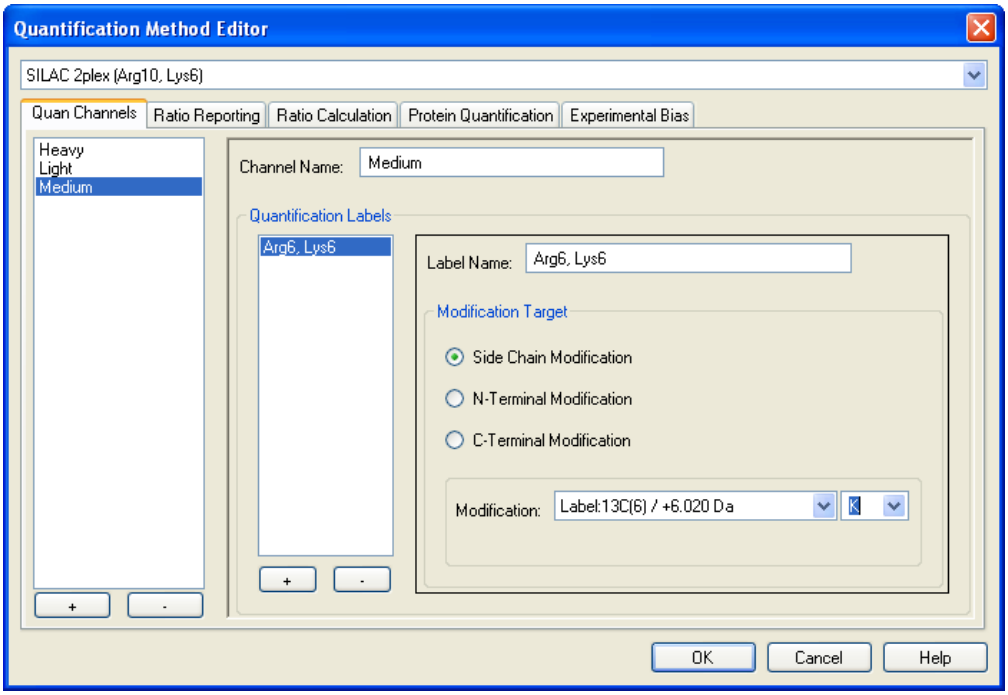
   A default quantification label of New *number* now appears in the Quantification Labels box and the Label Name box.

   To remove an existing quantification label, select the label in the Quantification Labels box and click **–** beneath the box.

6. To change the default quantification channel name, backspace over the name in the Label Name box and type the new name. The example in Figure 256 uses Arg6, Lys6.
7. In the Modification Target area, select the location of the label on the peptide:
   - Side Chain Modification: Indicates that the label occurs on a side chain.
   - N-Terminal Modification: Indicates that the label occurs on the N terminus.
   - C-Terminal Modification: Indicates that the label occurs on the C terminus.
8. From the Modification list, select the modification to label the amino acid with. This example shows Label:13C(6) / +6.020 Da.
9. From the list adjacent to the Modification list, select the abbreviation of the amino acid selected in the Quantification Labels box on which the modification should occur. In this example, K is selected.

   The completed Quan Channels page will resemble Figure 256.

   Figure 256. Completed Quan Channels page



10. Continue setting up the quantification method by following the instructions in "Setting Up Quantification Channels for Ratio Reporting" on page 495.

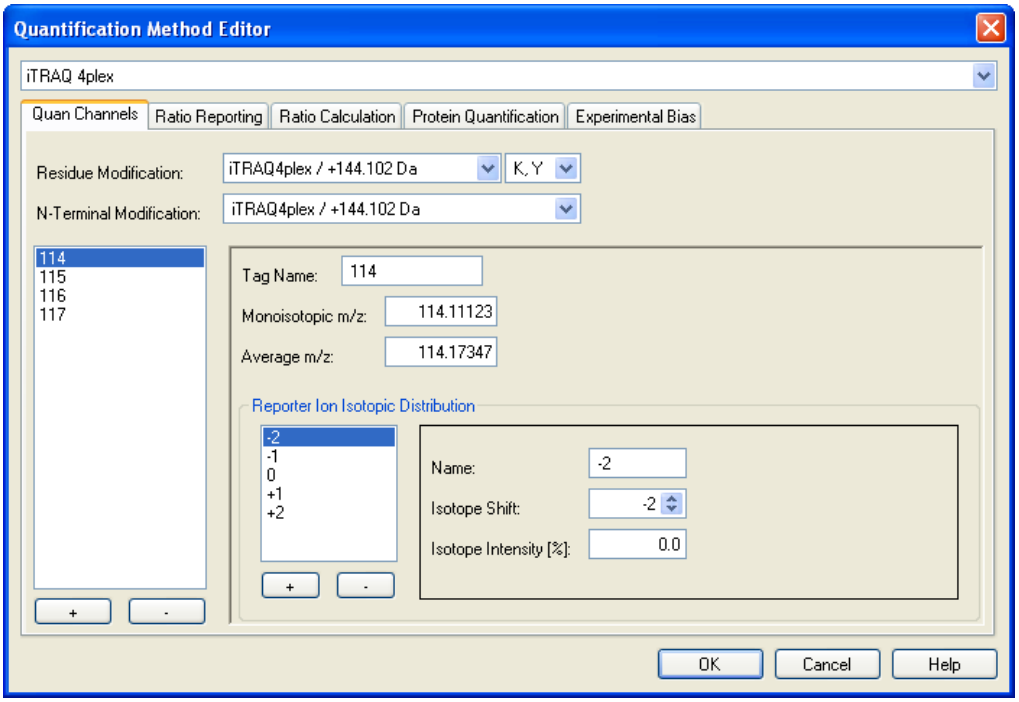
❖ To specify label modifications for reporter ion quantification

1. Click the **Quan Channels** tab of the Quantification Method Editor dialog box if it is not already selected.

2. In the top list, select the name of the method to use. For reporter ion quantification, you can select the following default methods when you initially set up a workflow and first access the Quantification Method Editor dialog box:

   - iTRAQ 4plex
   - iTRAQ 4plex (Thermo Scientific Instruments)
   - iTRAQ 8plex
   - iTRAQ 8plex (Thermo Scientific Instruments)
   - TMT 2plex
   - TMT 6plex

   Figure 257. Quan Channels page of the Quantification Method Editor dialog box for reporter ion quantification



3. From the Residue Modification list, select the label modification that would be found on the target amino acid residue. From the adjacent list, select the appropriate letter to indicate that the modification should occur on the indicated residue and will have an increased mass.

4. From the N-Terminal Modification list, select the label modification that would be found on the N terminus of each peptide.

   The left box of the Quan Channels page displays a list of mass tags, which are the fragmented labels.

5. To add a mass tag, click + beneath the list of mass tags in the box on the left.

   To remove a mass tag, select the mass tag you want to remove and click – beneath the list of mass tags.

6. When you add a mass tag or change the settings of an existing mass tag, do the following:

   a. In the Tag Name box, enter the name of the new mass tag if you do not want to use the default name.

   b. In the Monoisotopic m/z box, enter the monoisotopic mass-to-charge ratio of the new mass tag.

   c. In the Average m/z box, enter the average mass-to-charge ratio of the new mass tag.

   d. In the Reporter Ion Isotopic Distribution area, select the correction factor for the mass tags. Click + and – to add and delete correction factors.

      For information on these correction factors, see "Using Reporter Ion Isotopic Distribution Values to Correct for Impurities" on page 545.

      You must correct the purity of mass tags because of impurities in the tags themselves.

7. If you add a correction factor, do the following:

   a. In the Name box to the right of the list of correction factors, enter the name of the new correction factor.

      For the name, Thermo Fisher Scientific recommends that you use a plus (+) or a minus (–) symbol and the preferred shift number.

   b. In the Isotope Shift box, enter the isotope shift of the new correction factor.

      Isotope shift is a change in the spectral lines caused by different isotopes in an element. It often reflects impurities in the sample, and you must remove its corresponding mass-to-charge ratio from the calculations.

   c. In the Isotope Intensity [%] box, enter the isotope intensity of the new correction factor as a percentage.

      Isotope intensity is the intensity of the different isotopes in an element, often from impurities in the sample.

      Note The sum of the isotope intensities for each tag should add up to 100.

8. Continue setting up the quantification method by following the instructions in "Setting Up Quantification Channels for Ratio Reporting" on page 495.

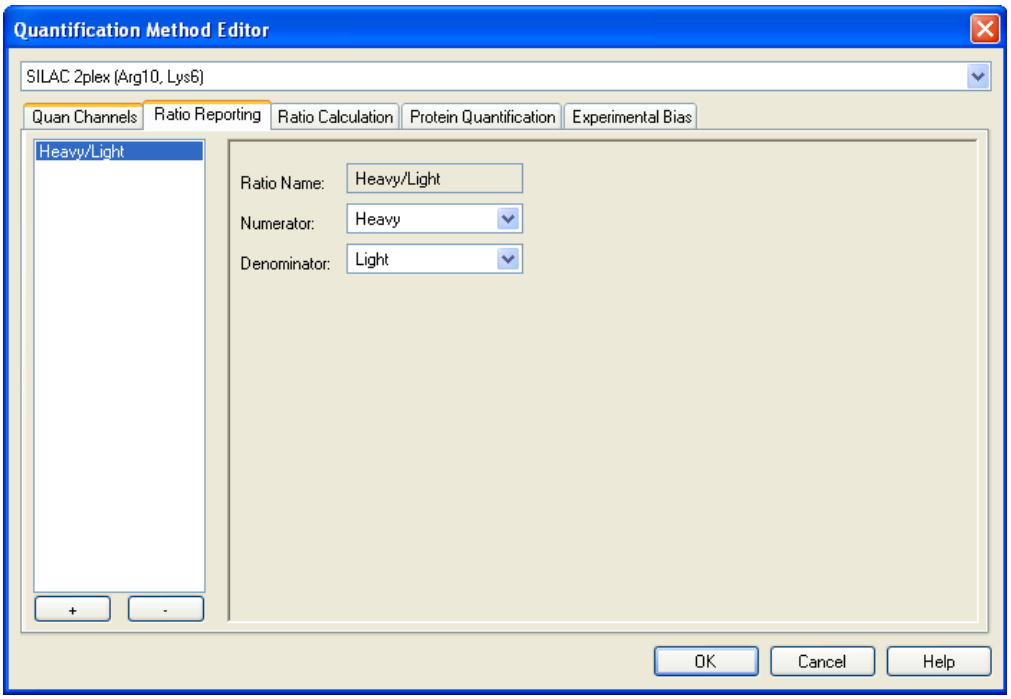
Setting Up Quantification Channels for Ratio Reporting

The Ratio Reporting page of the Quantification Method Editor dialog box specifies the names of the quantification channels (for precursor ion quantification) or mass tags (for reporter ion quantification) for the reporting of ratios that appear in the *Ratio* columns of the Proteins and Peptides pages.

❖ To set up the quantification channels for ratio reporting

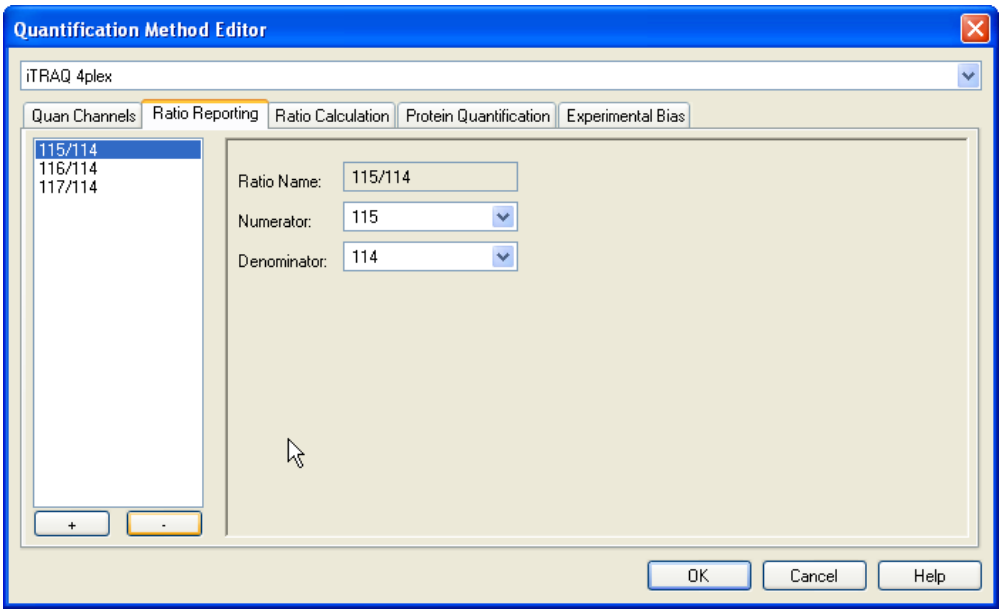
1. Click the **Ratio Reporting** tab, shown in Figure 258 for precursor ion quantification and in Figure 259 for reporter ion quantification.

   Figure 258. Ratio Reporting page of the Quantification Method Editor dialog box for precursor ion quantification



   In precursor ion quantification, the quantification ratios (left) box displays the ratio of the amino acids using heavy isotopes to the amino acids using normal isotopes.

Figure 259. Ratio Reporting page of the Quantification Method Editor dialog box for reporter ion quantification



In reporter ion quantification, the quantification ratios box to the left displays the name of the fragmented mass tag of a sample over the name of the mass tag of the reference sample.

2. To add any new quantification ratios, click **+** beneath the quantification ratios box.

   A new quantification ratio with the default name of New *number* appears in the quantification ratios pane.

   To remove the quantification ratio, select a quantification ratio and click **–** beneath the label type box.

3. If you added a quantification ratio, follow these steps.

   For precursor ion quantification:

   a. In the Numerator list, select the **Light** or **Heavy** label.

   b. In the Denominator list, select the **Light** or **Heavy** label that you did not select in the Numerator box.

   For reporter ion quantification:

   a. In the Numerator list, select the fragmented mass tag of the sample.

   b. In the Denominator list, select the name of the mass tag of the reference sample.

   You now see the specified numerator and denominator in the Ratio Name box, which is read-only.

4. Continue setting up the quantification method by following the instructions in "Setting Up the Ratio Calculation" on page 497.

Setting Up the Ratio Calculation

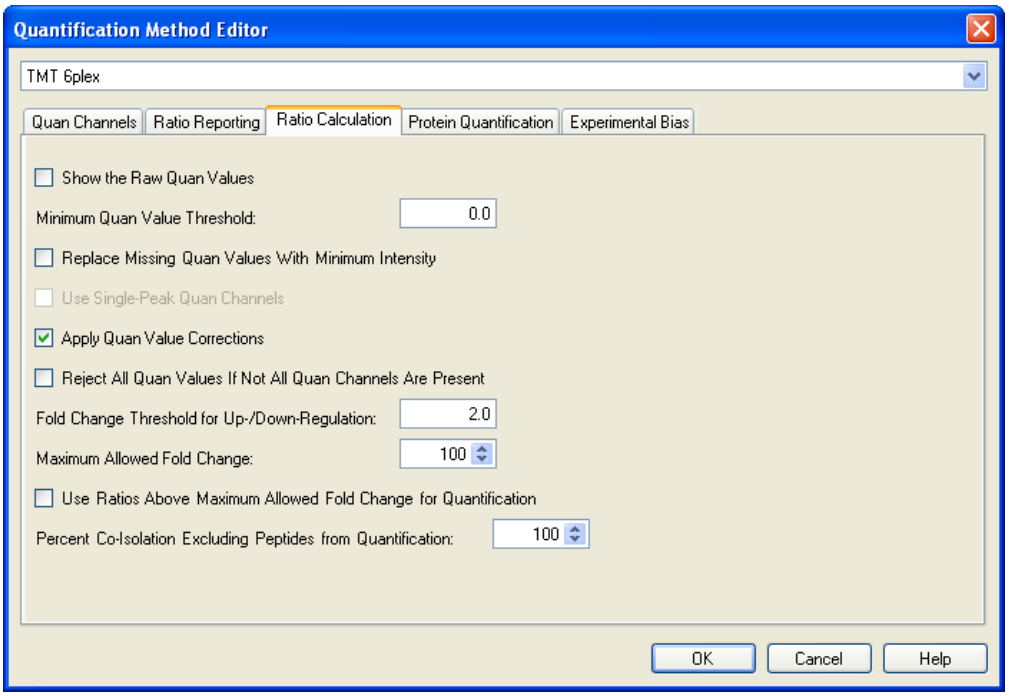
The Ratio Calculation page of the Quantification Method Editor dialog box controls how peptide and protein ratios are calculated from the raw quantification values of each quantification channel and how they are displayed on the Proteins and Peptides pages. For background information on the options available on this page, see "Missing Reporter Peaks in the Quantification Spectrum" on page 535.

❖ To set up the ratio calculation

1. Click the **Ratio Calculation** tab, shown in Figure 260.

   This page is the same for both precursor ion and reporter ion quantification.

   Figure 260. Ratio Calculation page of the Quantification Method Editor dialog box



2. To create additional columns in the results report that display the reporter ion intensities (or the corrected reporter ion intensities when you selected Apply Quan Value Corrections) for every peptide, select the **Show the Raw Quan Values** check box.

   By default, this option is clear.

3. To set all quantification values whose intensity falls below a specified threshold to zero, type the threshold in the Minimum Quan Value Threshold box.

   The default threshold value is 0.0.

4. When the ratio of the ion intensity of the peptide in a sample to the ion intensity of the peptide in the control sample is missing or is 0 and you want to replace it with the minimum ion intensity detected, select the **Replace Missing Quan Values With Minimum Intensity** check box.

   The Proteome Discoverer application searches for the minimum ion intensity that is detected on all quantification channels and uses it as a best guess for the detection limit. It then uses this minimum value instead of the missing quantification values. When you specify a value higher than the detected minimum value, the application uses the value that you specify instead. The Quantification Summary page lists the minimum quantification value detected and the value actually used for the calculations. For information on the Quantification Summary page, see "Summarizing the Quantification" on page 523.

   By default, this check box is clear.

5. When you are performing precursor ion quantification and want the Proteome Discoverer application to consider missing quantification channels or quantification channels with just one peak as valid quantification results in the ratio calculation, do the following:

   a. Select the **Use Single-Peak Quan Channels** check box on the Ratio Calculation page.

   b. Set the Single-Peak/Missing Channels Allowed parameter of the Precursor Ions Quantifier node to **1**.

      For more information on this parameter, see "Precursor Ions Quantifier Node" on page 699.

   By default, missing quantification channels or quantification channels with just one peak are not used for protein quantification. On the Peptides page, these peptides are marked "Excluded by Method" in the Quan Info column.

6. To apply the purity correction for the detected quantification values, select the **Apply Quan Value Corrections** check box.

   For reporter ion quantification, this option applies the correction for isotopic impurities. No such correction is currently available for precursor ion quantification. The application applies this purity correction after applying other settings that potentially change the quantification values.

   This option is selected by default.

7. To avoid using quantification values from any of the channels when one or more of the quantification channels has a detected intensity of zero, select the **Reject All Quan Values If Not All Quan Channels Are Present** check box.

   By default, this check box is clear.

8. To highlight a change in the ion intensity ratio (that is, the ratio of the ion intensity of the peptide in an experimental sample to the ion intensity of the peptide in the control sample) larger than *n* or smaller than 1/*n* in the results, specify *n* in the Fold Change Threshold for Up-/Down-Regulation box.

   The default is 2.0.

   For example, if you select 2 in the Fold Change Threshold for Up-/Down-Regulation box, the Proteome Discoverer application highlights those experimental results that are greater than twice as large (up-regulation) or less than half as large (down-regulation) as the control.

9. To exclude a peptide ion intensity ratio (that is, the ratio of the ion intensity of the peptide in a sample to the ion intensity of the peptide in the control sample) that exceeds a certain maximum, enter this maximum number in the Maximum Allowed Fold Change box.

   The minimum value is 1, and the maximum value is 100 000.

   The default is 100. With the default setting, calculated ratios above 100 are set to 100, and calculated ratios below 0.01 are set to 0.01.

   For example, if you set Maximum Allowed Fold Change to 10, the Proteome Discoverer application excludes any peptide ratios showing a greater than a ten-fold change in ion intensity for an experiment compared to the control.

10. To report larger ratios than you have indicated in the Maximum Allowed Fold Change box, select the **Use Ratios Above Maximum Allowed Fold Change for Quantification** check box.

    This option reports the quantification ratios based on the maximum values. Values greater than the value selected in the Maximum Allowed Fold Change box are replaced by the maximum or minimum value.

    By default, this check box is clear.

11. Continue setting up the quantification method by following the instructions in "Setting Peptide Parameters Used to Calculate Protein Ratios" on page 500.

The settings of the options on the Ratio Calculation page govern the appearance of the experimental results in the columns in the MSF report. The data can appear in the following colors:

- Pink: The experimental results are down-regulated.
- Blue: The experimental results are up-regulated.
- Red: The experimental results exceed the setting in the Maximum Allowed Fold Change box. These results are not used in calculations unless you select the Use Ratios Above Maximum Allowed Fold Change for Quantification option.

Setting Peptide Parameters Used to Calculate Protein Ratios

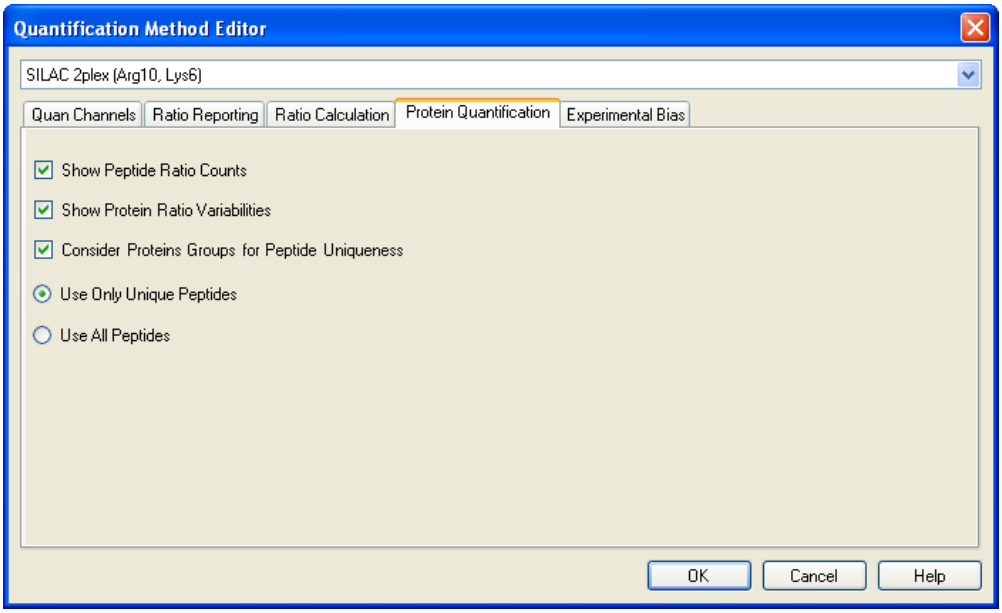
Use the Protein Quantification page of the Quantification Method Editor dialog box to set the peptide parameters for calculating protein ratios.

❖ To set the peptide parameters used to calculate protein ratios

1. Click the **Protein Quantification** tab, shown in Figure 261.

   This page is the same for both precursor ion and reporter ion quantification.

   Figure 261. Protein Quantification page of the Quantification Method Editor dialog box



2. If you want to display the number of peptide ratios that are used to calculate a protein ratio, select the **Show Peptide Ratio Counts** check box.

   The results appear in the Heavy/Light Count column of the Proteins page of the MSF report for precursor ion quantification and in the *Ratio* Count columns of the Proteins page for reporter ion quantification. For more information on ratio counts, see "Ratio Count" on page 561.

   This option is selected by default.

3. If you want to show the variability of the peptide ratios used to calculate the protein ratios, select the **Show Protein Ratio Variabilities** check box.

   The results appear in the Heavy/Light Variability [%] column of the Proteins page of the MSF report for precursor ion quantification and in the *Ratio* Variability [%] columns of the Proteins page for reporter ion quantification. For more information on protein variability, see "Ratio Variability" on page 561.

   This option is selected by default.

4. If you want to define peptide uniqueness on the basis of protein groups rather than on individual proteins, select the **Consider Proteins Groups for Peptide Uniqueness** check box.

   This option is selected by default.

5. Choose the type of peptides for the Proteome Discoverer application to use in the quantification:

   - (Default) Use Only Unique Peptides: Includes peptides that do not occur in other proteins.
   - Use All Peptides: Includes all detected peptides, whether or not they also occur in other proteins.

6. Continue setting up the quantification method by following the instructions in the next section, "Correcting Experimental Bias."

Correcting Experimental Bias

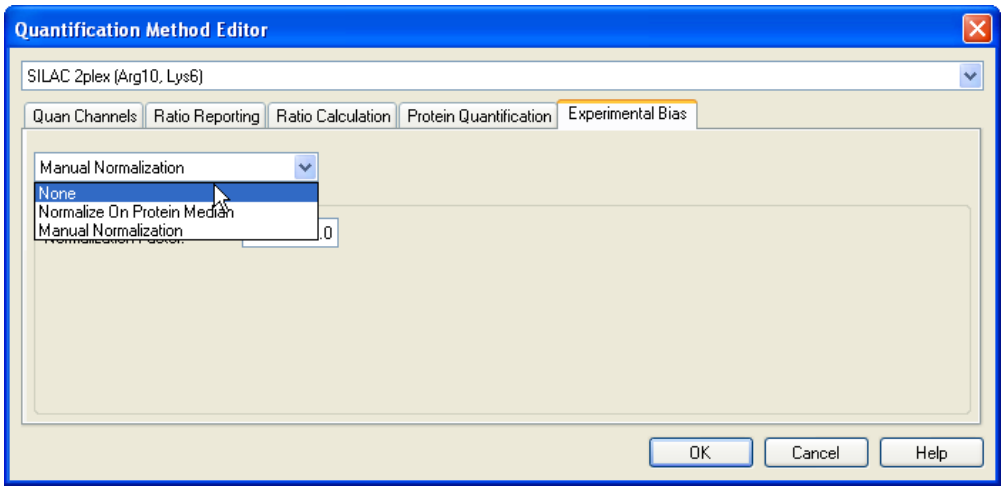
The purpose of the Experimental Bias page of the Quantification Method Editor dialog box is to correct experimental bias, which is the difference in the total observed protein abundance between two or more samples. Assuming that in real samples most of the proteins are not regulated, the intensity of the median protein in sample *x* should be the same as the intensity of the median protein in sample *y*. If it is not, it may indicate experimental bias caused by, for example, errors in pipetting or the determination of protein concentration in the mixed samples. You must correct for the difference. For best results, always enter a small normalization factor.

❖ To correct experimental bias

1. Click the **Experimental Bias** tab, shown in Figure 262.

   This page is the same for both precursor ion and reporter ion quantification.

   Figure 262. Experimental Bias page of the Quantification Method Editor dialog box



2. Select the normalization factor to apply from the list at the top of the page:
   - (Default) None: Performs no normalization.
   - Normalize on Protein Median: Normalizes all peptide ratios by the median protein ratio. The median protein ratio should be 1 after the normalization.
   - Manual Normalization: Specifies a user-defined normalization number.

   When you select the Normalize on Protein Median setting, the Minimum Protein Count box appears.

   When you select the Manual Normalization setting, the Normalization Factor box appears.
3. Do one of the following:
   - In the Minimum Protein Count box, which appears when you selected Normalization on Protein Median, enter the minimum number of proteins that must be observed to allow normalization.

   –or–

   - In the Normalization Factor box, which appears when you select Manual Normalization, enter the normalization factor.

   The default for the Minimum Protein Count option is 20, and the default for the Normalization Factor option is 1.0.

   Normalization cannot work if there are too few proteins in a sample.
4. Click **OK**.

Checking the Quantification Method

The Proteome Discoverer application checks the parameters that you have set for the quantification method. For reporter ion quantification, it verifies that the method has at least two channels. For precursor ion quantification, it checks for the following:

- At least one quantification channel
- At least one label for each quantification channel
- Unique label names in a channel
- The modification of each label applied to at least one amino acid, unless you chose None for a modification
- Each amino acid labeled only once in a channel. Labels must have an elemental composition defined.
- Each label mass used only once (label masses vary by at least 1 Da)

You cannot apply changes to a quantification method unless the method meets all these criteria.

Restoring Quantification Method Template Defaults

If you have altered one of the quantification method templates listed at the beginning of "Setting Up the Quantification Method" on page 485, you can restore the original template.

❖ To restore the original template

1. Choose **Administration > Maintain Quantification Methods**, or click on the **Maintain Quantification Methods** icon, .

   The Quantification Methods view opens, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

2. To open the Quantification Method Editor dialog box, click **Add** in the Quantification Methods view.

   The Create Quantification Method dialog box opens, as shown in Figure 268 on page 517.

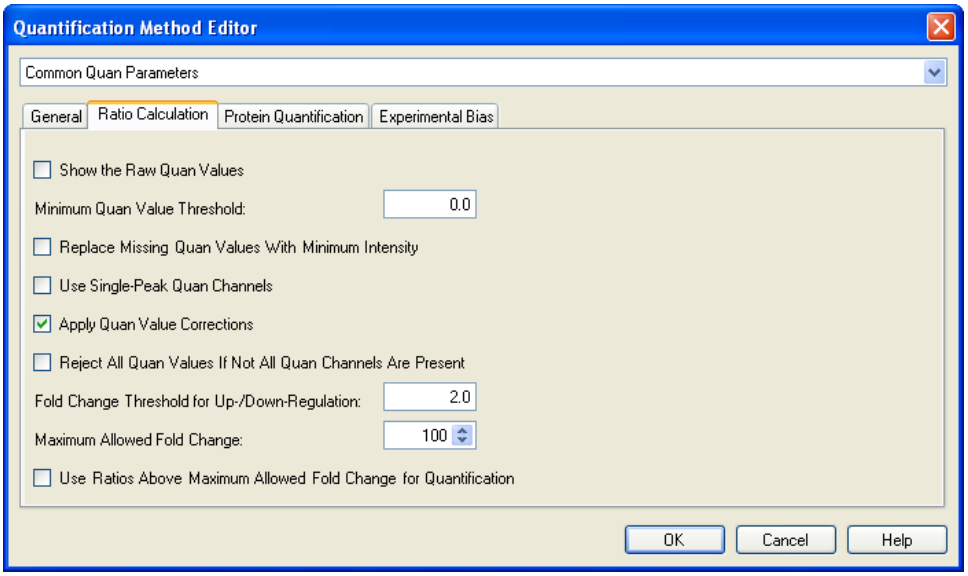
3. Select the appropriate template from the Create from Factory Defaults list.

4. Set up the quantification method according to the instructions in "Setting Up the Quantification Method" on page 485.

Setting Up the Quantification Method for Multiple Input Files

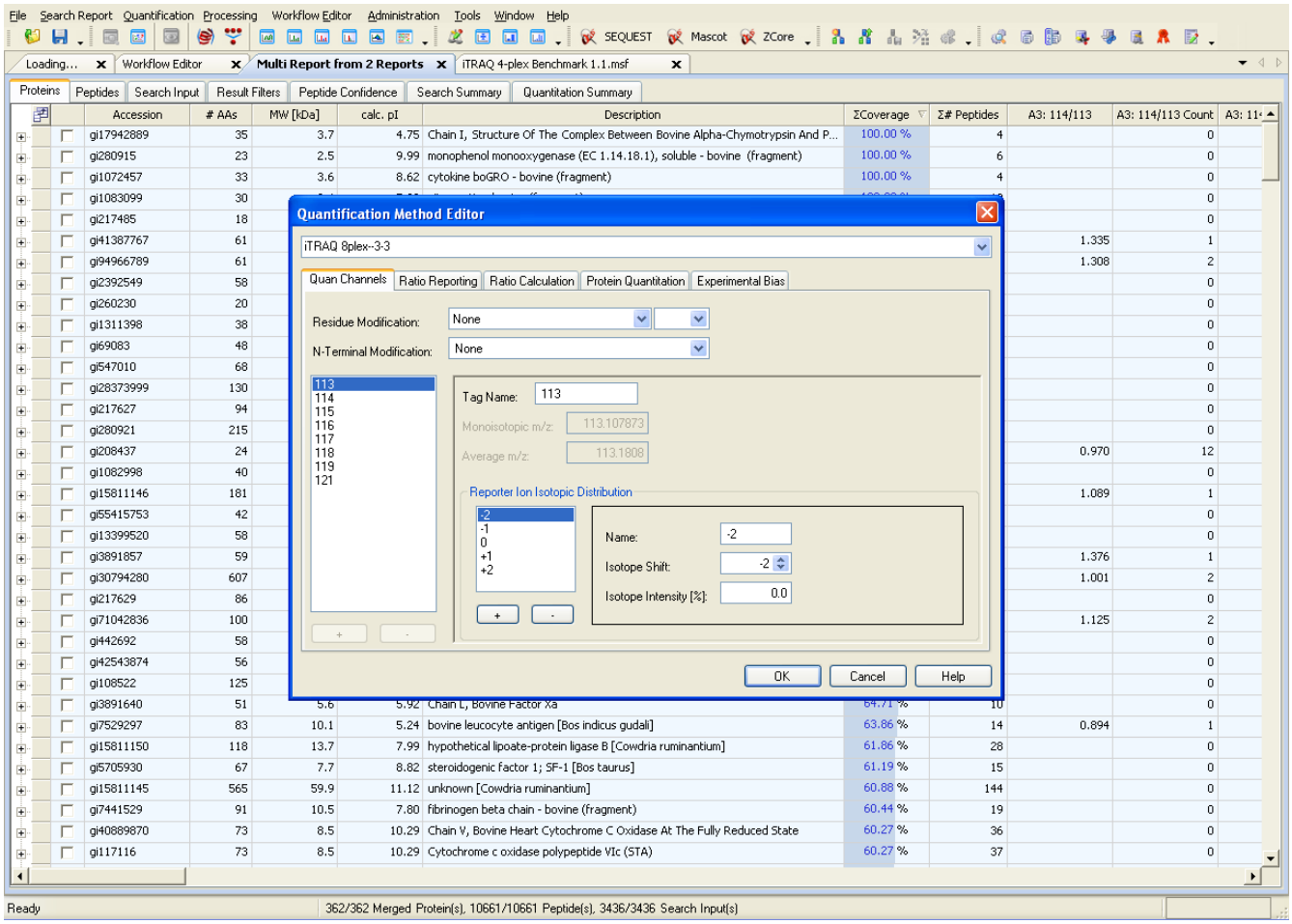
When you load multiple MSF files, you can apply the settings of the Ratio Calculation, Protein Quantification, and Experimental Bias pages of the Quantification Method Editor dialog box to all the loaded input files by selecting Common Quan Parameters from the list at the top of the dialog box, as shown in Figure 263. These pages contain the same options as those for single-file processing.

Figure 263. Quantification Method Editor dialog box for multiple input files



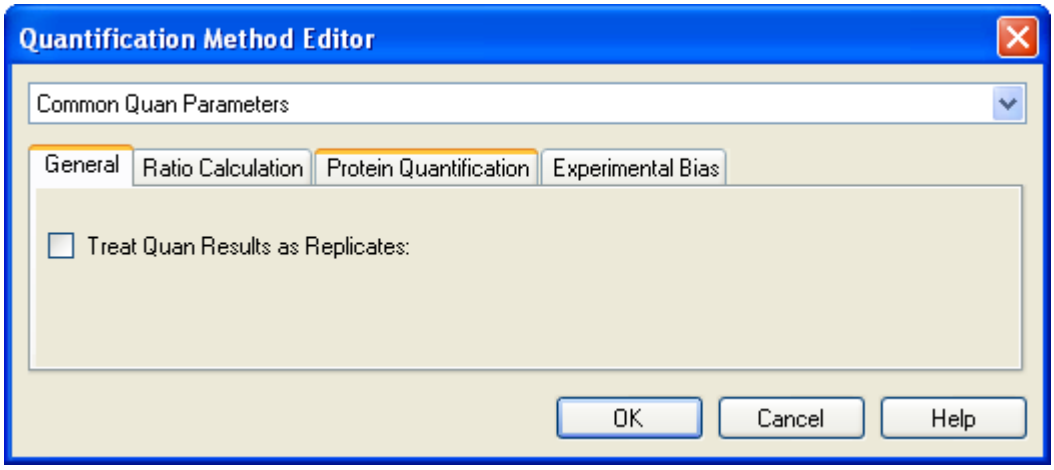
Although you cannot apply the settings of the Quan Channels and Ratio Reporting pages to multiple loaded MSF files, you can apply them to individual MSF files, as shown in Figure 264. You can access these two pages by selecting the individual MSF file from the list at the top of the dialog box.

Figure 264. Applying Quan Channel page settings to an individual MSF file when multiple MSF files are loaded



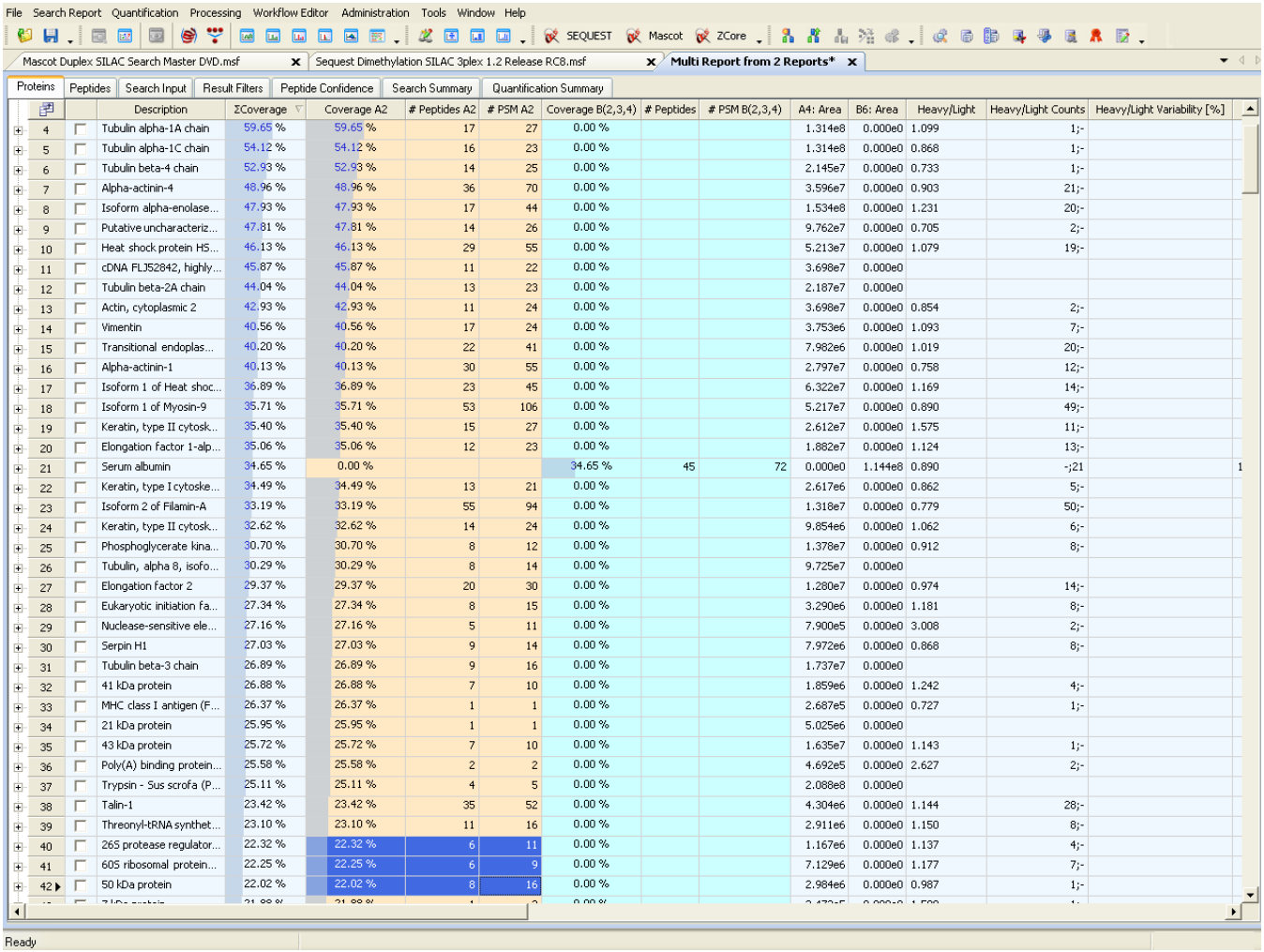
The Quantification Method Editor dialog box also includes a General page when multiple MSF files are loaded at the same time. It contains one option, Treat Quan Results as Replicates, as shown in Figure 265. This option treats protein-level quantification values with the same ratio names and the same quantification method as replicates (that is, the protein ratios of the individual files are averaged into a replicate ratio).

Figure 265. General page of the Quantification Method Editor dialog box



When you select Treat Quan Results as Replicates and click OK, the protein quantification data looks like the data in Figure 266.

Figure 266. Protein quantification data in replicate mode



		Description	ΣCoverage	Coverage A2	# Peptides A2	# PSM A2	Coverage B(2,3,4)	# Peptides	# PSM B(2,3,4)	A4: Area	B6: Area	Heavy/Light	Heavy/Light Counts	Heavy/Light Variability [%]
4		Tubulin alpha-1A chain	59.65 %	59.65 %	17	27	0.00 %			1.314e8	0.000e0	1.099	1;-	
5		Tubulin alpha-1C chain	54.12 %	54.12 %	16	23	0.00 %			1.314e8	0.000e0	0.868	1;-	
6		Tubulin beta-4 chain	52.93 %	52.93 %	14	25	0.00 %			2.145e7	0.000e0	0.733	1;-	
7		Alpha-actinin-4	48.96 %	48.96 %	36	70	0.00 %			3.596e7	0.000e0	0.903	21;-	
8		Isoform alpha-enolase...	47.93 %	47.93 %	17	44	0.00 %			1.534e8	0.000e0	1.231	20;-	
9		Putative uncharacteriz...	47.81 %	47.81 %	14	26	0.00 %			9.762e7	0.000e0	0.705	2;-	
10		Heat shock protein HS...	46.13 %	46.13 %	29	55	0.00 %			5.213e7	0.000e0	1.079	19;-	
11		cDNA FLJ52842, highly...	45.87 %	45.87 %	11	22	0.00 %			3.698e7	0.000e0			
12		Tubulin beta-2A chain	44.04 %	44.04 %	13	23	0.00 %			2.187e7	0.000e0			
13		Actin, cytoplasmic 2	42.93 %	42.93 %	11	24	0.00 %			3.698e7	0.000e0	0.854	2;-	
14		Vimentin	40.56 %	40.56 %	17	24	0.00 %			3.753e6	0.000e0	1.093	7;-	
15		Transitional endoplas...	40.20 %	40.20 %	22	41	0.00 %			7.982e6	0.000e0	1.019	20;-	
16		Alpha-actinin-1	40.13 %	40.13 %	30	55	0.00 %			2.797e7	0.000e0	0.758	12;-	
17		Isoform 1 of Heat shoc...	36.89 %	36.89 %	23	45	0.00 %			6.322e7	0.000e0	1.169	14;-	
18		Isoform 1 of Myosin-9	35.71 %	35.71 %	53	106	0.00 %			5.217e7	0.000e0	0.890	49;-	
19		Keratin, type II cytosk...	35.40 %	35.40 %	15	27	0.00 %			2.612e7	0.000e0	1.575	11;-	
20		Elongation factor 1-alp...	35.06 %	35.06 %	12	23	0.00 %			1.882e7	0.000e0	1.124	13;-	
21		Serum albumin	34.65 %	0.00 %			34.65 %	45	72	0.000e0	1.144e8	0.890	-;21	1
22		Keratin, type I cytoske...	34.49 %	34.49 %	13	21	0.00 %			2.617e6	0.000e0	0.862	5;-	
23		Isoform 2 of Filamin-A	33.19 %	33.19 %	55	94	0.00 %			1.318e7	0.000e0	0.779	50;-	
24		Keratin, type II cytosk...	32.62 %	32.62 %	14	24	0.00 %			9.854e6	0.000e0	1.062	6;-	
25		Phosphoglycerate kina...	30.70 %	30.70 %	8	12	0.00 %			1.378e7	0.000e0	0.912	8;-	
26		Tubulin, alpha 8, isofo...	30.29 %	30.29 %	8	14	0.00 %			9.725e7	0.000e0			
27		Elongation factor 2	29.37 %	29.37 %	20	30	0.00 %			1.280e7	0.000e0	0.974	14;-	
28		Eukaryotic initiation fa...	27.34 %	27.34 %	8	15	0.00 %			3.290e6	0.000e0	1.181	8;-	
29		Nuclease-sensitive ele...	27.16 %	27.16 %	5	11	0.00 %			7.900e5	0.000e0	3.008	2;-	
30		Serpin H1	27.03 %	27.03 %	9	14	0.00 %			7.972e6	0.000e0	0.868	8;-	
31		Tubulin beta-3 chain	26.89 %	26.89 %	9	16	0.00 %			1.737e7	0.000e0			
32		41 kDa protein	26.88 %	26.88 %	7	10	0.00 %			1.859e6	0.000e0	1.242	4;-	
33		MHC class I antigen (F...	26.37 %	26.37 %	1	1	0.00 %			2.687e5	0.000e0	0.727	1;-	
34		21 kDa protein	25.95 %	25.95 %	1	1	0.00 %			5.025e6	0.000e0			
35		43 kDa protein	25.72 %	25.72 %	7	10	0.00 %			1.635e7	0.000e0	1.143	1;-	
36		Poly(A) binding protein...	25.58 %	25.58 %	2	2	0.00 %			4.692e5	0.000e0	2.627	2;-	
37		Trypsin - Sus scrofa (P...	25.11 %	25.11 %	4	5	0.00 %			2.088e8	0.000e0			
38		Talin-1	23.42 %	23.42 %	35	52	0.00 %			4.304e6	0.000e0	1.144	28;-	
39		Threonyl-tRNA synthet...	23.10 %	23.10 %	11	16	0.00 %			2.911e6	0.000e0	1.150	8;-	
40		265 protease regulator...	22.32 %	22.32 %	6	11	0.00 %			1.167e6	0.000e0	1.137	4;-	
41		605 ribosomal protein...	22.25 %	22.25 %	6	9	0.00 %			7.129e6	0.000e0	1.177	7;-	
42		50 kDa protein	22.02 %	22.02 %	8	16	0.00 %			2.984e6	0.000e0	0.987	1;-	

There is only one ratio column for each specified ratio, and the *Ratio* Counts columns show the number of peptides used from every single MSF file for calculating the individual protein ratios for each file. These individual protein ratios are then averaged, and the average is displayed in the ratio columns for the proteins.

Quantification Method Editor Dialog Box Parameters

This section describes the parameters available on the five pages of the Quantification Method Editor dialog box.

- Quan Channels Page Parameters
- Ratio Reporting Page Parameters
- Ratio Calculation Page Parameters
- Protein Quantification Page Parameters
- Experimental Bias Page Parameters

Quan Channels Page Parameters

Table 79 lists the parameters on the Quan Channels page of the Quantification Method Editor dialog box for precursor ion quantification.

Table 79. Quan Channels page parameters for precursor ion quantification (Sheet 1 of 2)

Parameter	Description
Top box	Specifies the quantification method to use. For precursor ion quantification, you can select the following default methods when you initially set up a workflow and first access the Quantification Method Editor dialog box: • ^{18}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O. • Dimethylation 3plex: Chemically adds isotopically labeled dimethyl groups to the N-terminus and to the ε-amino group of lysine. • SILAC 2plex (Arg10, Lys6): Uses arginine 10 and lysine 6. • SILAC 2plex (Arg10, Lys8): Uses arginine 10 and lysine 8. • SILAC 2plex (Ile6): Uses isoleucine 6. • SILAC 3plex (Arg6, Lys4|Arg10, Lys8): Uses arginine 10 and lysine 8 for "heavy" labels and arginine 6 and lysine 4 for "medium" labels. • SILAC 3plex (Arg6, Lys6|Arg10, Lys8): Uses arginine 10 and lysine 8 for "heavy" labels and arginine 6 and lysine 6 for "medium" labels. For more information on these methods, see "Performing Precursor Ion Quantification" on page 469. For multiconsensus reports containing quantification results, the list at the top of the dialog box contains the Common Quan Parameters option, which applies the quantification parameters that you specify in the dialog box to all the result files at once. You can also use this list to select the individual MSF file quantification parameters.
Quantification channels (left) pane	Lists the quantification channels available. Heavy, Light, and Medium (3plex methods only) are available by default.
+	Adds a new quantification channel to the list of quantification channels.
–	Removes a quantification channel from the list of quantification channels.
Channel Name	Displays the name of the quantification channel that you selected in the quantification channels pane.

Table 79. Quan Channels page parameters for precursor ion quantification (Sheet 2 of 2)

Parameter	Description
Quantification Labels	Lists the amino acid labels assigned to the selected label type.
+	Adds a new label name to the list of amino acid label names.
–	Removes a label name from the list of amino acid label names.
Label Name	Specifies the name of the selected quantification label.
Modification Target	Specifies the location of the label on the amino acid: • Side Chain Modification: Indicates that the modification occurs on a side chain. • N-Terminal Modification: Indicates that the modification occurs on the N terminal. • C-Terminal Modification: Indicates that the modification occurs on the C terminal.
Modification	Specifies the modification to label the amino acid with and the abbreviation of the amino acid selected in the Quantification Labels box on which the modification should occur.

Table 80 lists the parameters on the Quan Channels page of the Quantification Method Editor dialog box for reporter ion quantification.

Table 80. Quan Channels page parameters for reporter ion quantification (Sheet 1 of 2)

Parameter	Description
Top box	Specifies the quantification method to use. For reporter ion quantification, you can select the following default methods when you initially set up a workflow and first access the Quantification Method Editor dialog box: • iTRAQ 4plex • iTRAQ 4plex (Thermo Scientific Instruments) • iTRAQ 8plex • iTRAQ 8plex (Thermo Scientific Instruments) • TMT 2plex • TMT 6plex For more information on these methods, see "Performing Reporter Ion Quantification" on page 476. The two methods labeled "Thermo Scientific Instruments" have purity corrections optimized for the way Thermo Scientific mass spectrometers process samples and produce data. For multiconsensus reports containing quantification results, the list at the top of the dialog box contains the Common Quan Parameters option, which applies the quantification parameters that you specify in the dialog box to all the result files at once. You can also use this list to select the individual MSF file quantification parameters.
Residue Modification	Specifies the label modification that would be found on the target amino acid residue.
N-Terminal Modification	Specifies the label modification that would be found on the N terminus of each peptide.
Mass tags (left) pane	Lists the names of the mass tags available, which are the fragmented labels.
+	Adds a new mass tag to the mass tags pane and to the Tag Name box. You can rename this mass tag name.
–	Deletes the selected mass tag name from the mass tags pane.
Tag Name	Displays the name of the selected or new mass tag.
Monoisotopic m/z	Displays the monoisotopic mass-to-charge ratio of the selected or new mass tag.

Table 80. Quan Channels page parameters for reporter ion quantification (Sheet 2 of 2)

Parameter	Description
Average m/z	Specifies the average mass-to-charge ratio of the selected or new mass tag.
Reporter Ion Isotopic Distribution	Specifies the correction factors for the mass tags.
+	Adds the name of a correction factor to the Reporter Ion Isotopic Distribution box and to the Name box in the Reporter Ion Isotopic Distribution area.
–	Deletes a correction factor for the mass tags.
Name	Specifies the name of the new correction factor. For the name, Thermo Fisher Scientific recommends that you use a plus (+) or a minus (–) symbol and the appropriate shift number.
Isotope Shift	Specifies the isotope shift of the new correction factor. Isotope shift is a change in the spectral lines caused by different isotopes in an element. It often reflects impurities in the sample, and its corresponding mass-to-charge ratio must be removed from the calculations.
Isotope Intensity [%]	Specifies the isotope intensity of the new correction factor as a percentage.

Ratio Reporting Page Parameters

Table 81 lists the parameters on the Ratio Reporting page of the Quantification Method Editor dialog box for precursor ion quantification.

Table 81. Ratio Reporting page parameters for precursor ion quantification

Parameter	Description
Quantification ratios (left) pane	Displays the ratio of the amino acids using heavy isotopes to the amino acids using normal isotopes. If you use a Medium channel, the normal isotopes might also be labeled.
+	Adds a new ratio.
–	Deletes the selected ratio.
Ratio Name	Displays the ratio selected in the Numerator and Denominator lists. This field is read-only.
Numerator	Selects the channels used in the numerator for ratio calculation.
Denominator	Selects the channels used in the denominator for ratio calculation.

Table 82 lists the parameters on the Ratio Reporting page of the Quantification Method Editor dialog box for reporter ion quantification.

Table 82. Ratio Reporting page parameters for reporter ion quantification

Parameter	Description
Mass tags (left) pane	Displays the name of the fragmented mass tag of a sample over the name of the mass tag of the reference sample.
+	Adds a mass tag.
–	Deletes the selected mass tag.
Ratio Name	Displays the ratio selected in the Numerator and Denominator lists. This field is read-only.
Numerator	Displays the name of the fragmented mass tag of the sample.
Denominator	Displays the name of the mass tag of the reference sample.

Ratio Calculation Page Parameters

Table 83 lists the parameters on the Ratio Calculation page of the Quantification Method Editor dialog box.

Table 83. Ratio Calculation page parameters (Sheet 1 of 3)

Parameter	Description
Show the Raw Quan Values	Creates additional columns in the results report that display the reporter ion intensities (or the corrected reporter ion intensities if you selected the Apply Quan Value Corrections option for every peptide. By default, this check box is clear.
Minimum Quan Value Threshold	Sets all quantification values whose intensity falls below a specified threshold to zero. The default threshold value is 0.0.
Replace Missing Quan Values With Minimum Intensity	Uses the minimum intensity detected as a best guess for the detection limit if the ratio of the intensity of the peptide in a sample to the intensity of the peptide in the control sample is missing or 0. By default, this check box is clear.

Table 83. Ratio Calculation page parameters (Sheet 2 of 3)

Parameter	Description
Use Single-Peak Quan Channels	Reports the ratio of the existing quantification channels (for example, the heavy and light) even if one channel (for example, medium) is missing. The Use Single-Peak Quan Channels check box is available only for precursor ion quantification. When you select the Use Single-Peak Quan Channels check box, you can set the Single-Peak/Missing Channels Allowed parameter of the Precursor Ions Quantifier node to a value other than 1. For more information on this parameter, see "Precursor Ions Quantifier Node" on page 699. By default, missing quantification channels or quantification channels with just one peak are not used for protein quantification. On the Peptides page, these peptides are marked "Excluded by Method" in the Quan Info column.
Apply Quan Value Corrections	Applies the correction for isotopic impurities to the detected quantification values. The Proteome Discoverer application applies this purity correction after applying other settings that potentially change the quantification values. The Apply Quan Value Corrections check box is available only for reporter ion quantification. This value is selected by default.
Reject All Quan Values If Not All Quan Channels Are Present	Sets all quantification values from all channels to zero if one or more of the quantification channels has a detected intensity of zero. By default, this check box is clear.
Fold Change Threshold for Up/Down Regulation	Specifies an ion intensity ratio, *n*, that the Proteome Discoverer application highlights in the results if it is larger than *n* or smaller than 1/*n*. The default is 2.0.
Maximum Allowed Fold Change	Reports the quantification ratios based on the maximum values. Values greater than the value selected by this option are replaced by the maximum or minimum value. The minimum value is 1, and the maximum value is 100 000. The default is 100. With the default setting, ratios above 100 are set to 100, and ratios below 0.01 are set to 0.01.

Table 83. Ratio Calculation page parameters (Sheet 3 of 3)

Parameter	Description
Use Ratios Above Maximum Allowed Fold Change for Quantification	Reports larger ratios than those indicated in the Maximum Allowed Fold Change box. This option ignores the setting of the Maximum Allowed Fold Change option and reports the exact calculation factor. By default, this check box is clear.
Percent Co-Isolation Excluding Peptides from Quantification	Specifies a threshold for the amount of interference allowed by co-isolation. You can select a value between 0 percent and 100 percent. The default value is 100 percent, which means that no PSM is excluded. This parameter is only available for reporter ion quantification. For more information on co-isolation, see "Excluding Peptides with High Levels of Co-Isolation" on page 547.

Protein Quantification Page Parameters

Table 84 lists the parameters on the Protein Quantification page of the Quantification Method Editor dialog box.

Table 84. Protein Quantification page parameters (Sheet 1 of 2)

Parameter	Description
Show Peptide Ratio Counts	Displays the number of peptide ratios that are used to calculate a protein ratio in the Heavy/Light (or Medium/Light) Count column of the Proteins page of the MSF report for precursor ion quantification and in the *Ratio* Count columns of the Proteins page for reporter ion quantification. This option is selected by default.
Show Protein Ratio Variabilities	Displays the variability of the protein ratios in the Heavy/Light (or Medium/Light) Variability [%] column of the Proteins page of the MSF report for precursor ion quantification and in the *Ratio* Variability [%] columns of the Proteins page for precursor ion or reporter ion quantification. For more information on protein variability, see "Ratio Variability" on page 561. This option is selected by default.

Table 84. Protein Quantification page parameters (Sheet 2 of 2)

Parameter	Description
Consider Proteins Groups for Peptide Uniqueness	Defines peptide uniqueness on the basis of protein groups rather than individual proteins. This option is selected by default.
Use Only Unique Peptides/Use All Peptides	Specifies the type of peptides for the Proteome Discoverer application to use in the quantification: • (Default) Use Only Unique Peptides: Includes peptides that do not occur in other proteins. • Use All Peptides: Includes all peptides, whether or not they occur in other proteins.

Experimental Bias Page Parameters

Table 85 lists the parameters on the Experimental Bias page of the Quantification Method Editor dialog box.

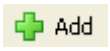
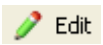
Table 85. Experimental Bias page parameters

Parameter	Description
Normalize on Protein Median/Manual Normalization/None	Specifies the normalization factor to apply: • Normalize on Protein Median: For each channel, normalizes all peptide ratios by the median peptide ratio. The median protein ratio should be 1. • Manual Normalization: Specifies a user-defined normalization number. • (Default) None: Performs no normalization.
Minimum Protein Count	Specifies the minimum number of proteins that must be observed to allow normalization. This option becomes available when you select Normalize on Protein Median. The default is 20.
Normalization Factor	Specifies the normalization factor. This option becomes available when you select Manual Normalization. The default is 1.0.

Quantification Methods View Parameters

Table 86 describes the parameters in the Quantification Methods view.

Table 86. Quantification Methods view parameters

Parameter	Description
	Opens the Create Quantification Method dialog box, shown in Figure 268 on page 517, so that you can specify how to create a new quantification method.
	Deletes the selected quantification method.
	Opens the Quantification Method Editor dialog box so that you can change the selected quantification method.
	Imports a new quantification method.
	Exports an existing quantification method.
Status	Indicates whether the quantification method is valid for use in quantification. • A green check mark means that the quantification method is valid and can be used for quantification. • An exclamation point in a yellow triangle means that the quantification method is not valid. Double-click this mark to view a message that describes the error and tells you how to fix it.
Method Name	Displays the name of the available quantification methods.
Description	Displays a brief description of the available quantification methods.
Is Active	Indicates with a check mark that a quantification method is available.

Adding a Quantification Method

You can use the following procedure to add a quantification method. You can also use it to access the quantification methods without loading an MSF file.

❖ To add a quantification method

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon,  .

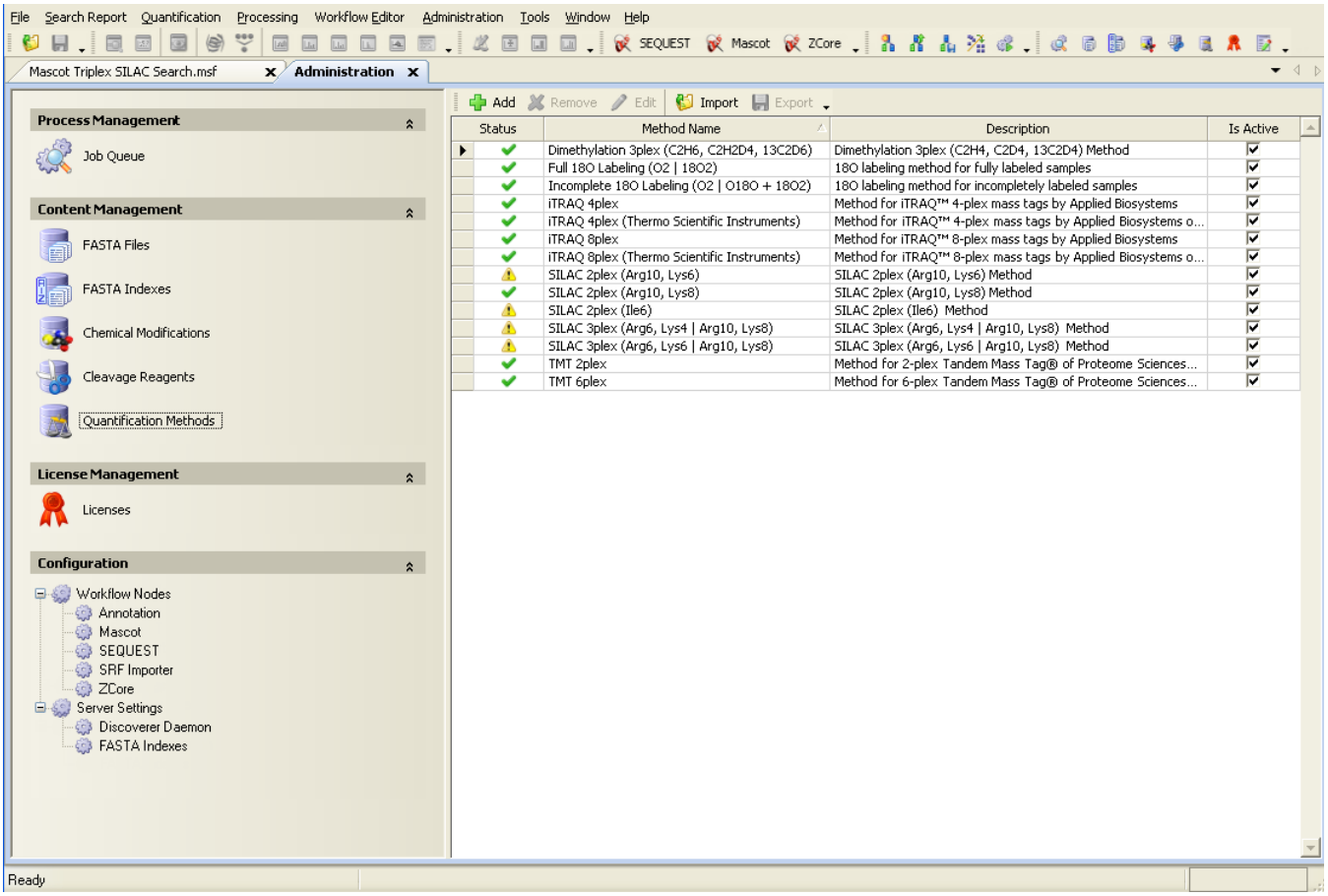
   The Quantification Methods view opens, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

   The Status column indicates whether the quantification method is valid for use in quantification:

   - A green check mark means that the quantification method is valid and can be used for quantification.
   - An exclamation point in a yellow triangle means that the quantification method is not valid. Double-click this mark to view a message that describes the error and provides information on how to fix it.

   Figure 267 provides examples of these symbols in the Status column.

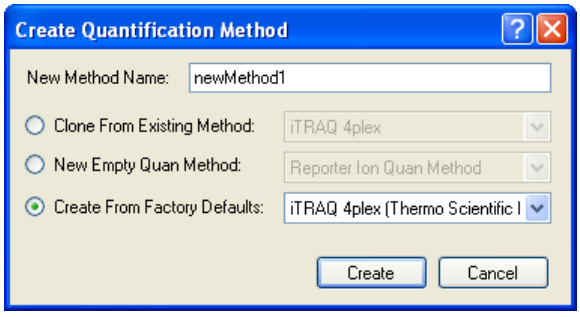
Figure 267. Method validity symbols in the Quantification Methods view



2. Click **Add**.

   The Create Quantification Method dialog box now appears, as shown in Figure 268.

   Figure 268. Create Quantification Method dialog box



3. In the New Method Name box, type the name of the quantification method that you want to create.

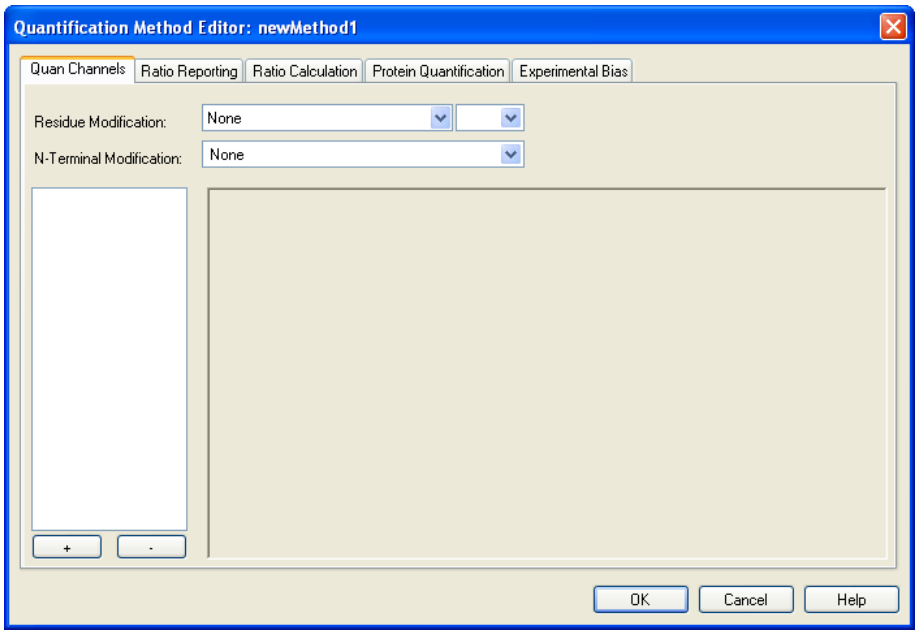
4. Select one of the following methods of creating a quantification method:

   - Clone From Existing Method: Uses the same settings as those of the existing quantification method that you select from the list. The list of methods is the same as that given at the beginning of "Setting Up the Quantification Method" on page 485.
   - New Empty Quan Method: Uses one of the following templates so that you can build a new processing method from scratch:
     - Reporter Ion Quan Method: Provides a template for reporter ion quantification.
     - Precursor Ion Quan Method: Provides a template for precursor ion quantification.
   - Create From Factory Defaults: Creates a new method using the same settings from one of the default settings that appear when the Proteome Discoverer application is newly installed.

5. Click **Create**.

   The Quantification Method Editor dialog box appears, as shown in Figure 257 on page 493 through Figure 262 on page 501. The Quan Channels page and the Ratio Reporting page are blank if you selected the New Empty Quan Method option. In this case, the Quan Channels page resembles Figure 269.

   Figure 269. Empty quantification method template



6. To specify the parameters of the new quantification method, follow the procedure given in "Setting Up the Quantification Method" on page 485.

Create Quantification Method Dialog Box Parameters

Table 87 describes the parameters in the Create Quantification Method dialog box.

Table 87. Create Quantification Method dialog box parameters

Parameter	Description
New Method Name	Allows you to specifies the name of the method that you want to create.
Clone From Existing Method	Uses the same settings as those of the existing quantification method selected from the list. The list of methods is the same as that given at the beginning of "Setting Up the Quantification Method" on page 485.
New Empty Quan Method	Uses one of the following templates so that you can build a new processing method from scratch: • Reporter Ion Quan Method: Provides a template for reporter quantification. • Precursor Ion Quan Method: Provides a template for precursor ion quantification.
Create From Factory Defaults	Uses the same settings from one of the default settings that appear when the Proteome Discoverer application is newly installed.
Create	Opens the Quantification Method Editor dialog box using the method that you specified.

Changing a Quantification Method

After you perform quantification, you can change the quantification method of the current report. You can add new quantification methods by copying an existing method and editing it. You can also activate and deactivate methods that you want visible or hidden when setting up a quantification workflow. However, you cannot define mass tags or labels as you can when setting up the initial quantification method, because they have already been measured.

You can access the quantification methods without loading an MSF file by using Administration > Maintain Quantification Methods (see "Adding a Quantification Method" on page 516), but if you want to save any changes to the quantification method in a report, you must first open that report and use Quantification > Edit Quantification Method.

❖ To change a quantification method

1. Select the MSF report whose quantification method you want to change.
2. Do one of the following:

   - Choose **Quantification > Edit Quantification Method**, or click the **Edit Quantification Method** icon,  , either on the toolbar or on the Administration page.

     Note To access the Edit Quantification Method command, you must first run a workflow that uses the Reporter Ions Quantifier node or the Precursor Ions Quantifier node.

     The Quantification Method Editor dialog box appears, as shown in Figure 257 on page 493 through Figure 262 on page 501.

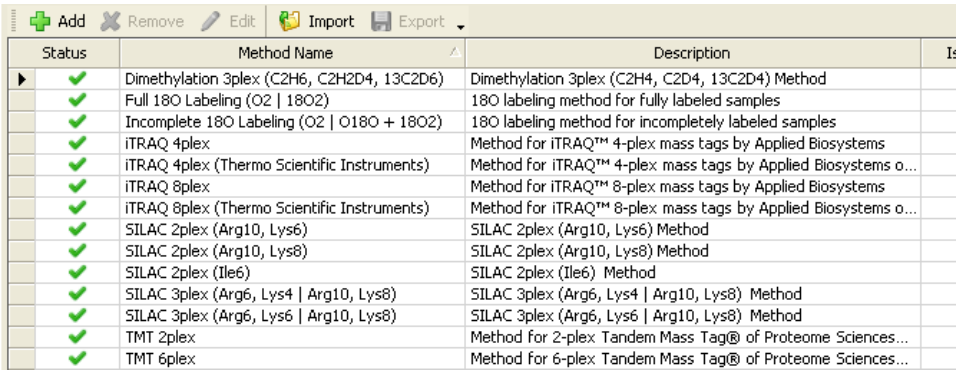
   –or–

   - Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon,  .

     The Quantification Methods view appears, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

     Then, either double-click the row for the appropriate method in the Method Name or Description column, or click the column to the left of Method Name for the method, as shown in Figure 270, and click **Edit**.

Figure 270. Selecting the method to edit



The Quantification Method Editor dialog box appears, as shown in Figure 257 on page 493 through Figure 262 on page 501.

3. Follow the procedure given in "Setting Up the Quantification Method" on page 485.

The Proteome Discoverer application checks the parameters that you have changed to be sure that they conform to the guidelines given in "Checking the Quantification Method" on page 502. It does not apply the changes to a quantification method unless the method meets all these criteria.

The changes that you make to a quantification method only affect the method in the selected results report.

Removing a Quantification Method

You can delete a quantification method if it is no longer useful, or make a quantification method temporarily unavailable to new workflows.

❖ To remove a quantification method

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, .

   The Quantification Methods view opens, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Click the box to the left of the method that you want to remove.

   The Remove icon, Remove, now becomes available.

3. Click Remove.

4. In the Delete Methods dialog box, click **OK**.

❖ **To deactivate a quantification method**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon,  .

   The Quantification Methods view opens, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Clear the check box in the Is Active column on the line containing the quantification method that you want to render inactive.

To make the quantification method active again, select the same check box.

Importing a Quantification Method

You can import a new quantification method from another computer.

❖ **To import a quantification method**

1. Choose **Administration > Maintain Quantification Methods** or click the **Maintain Quantification Methods** icon,  , either on the toolbar or on the Administration page.

   The Quantification Methods view opens, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Click  .

3. In the Import Quan Method dialog box, select the .method file containing the method that you want to import, and click **Open**.

   - If the new method is valid, the Quantification Method Editor dialog box opens, showing the new method.
   - If the new method is not valid, a message box appears that describes the error.

4. If the new method is valid, click **OK** in the Quantification Method Editor dialog box.

5. Change the name of the imported quantification method by changing it in the table of the Quan Method Manager.

Exporting a Quantification Method

You can save a quantification method to use on another computer.

❖ **To export a quantification method**

1. Choose **Administration > Maintain Quantification Methods** or click the **Maintain Quantification Method**s icon, , either on the toolbar or on the Administration page.

   The Quantification Methods view opens, as shown in Figure 253 on page 487. It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Select the method that you want to export in the Quantification Methods view by clicking in the leftmost column.

3. Click Export .

4. In the Export Quan Method dialog box, select the name of the .method file containing the quantification method to be exported, and click **Save**.

Summarizing the Quantification

The Quantification Summary page, shown in Figure 271 for precursor ion quantification and in Figure 272 for reporter ion quantification, summarizes the settings that you chose for the Precursor Ions Quantifier node or the Reporter Ions Quantifier node in the parameters pane of the Workflow Editor. It also shows the settings that you chose on the pages of the Quantification Method Editor for precursor ion and reporter ion quantification.

You must conduct a search with a workflow that includes a quantification node for this page to appear.

Figure 271. Quantification Summary page for precursor ion quantification

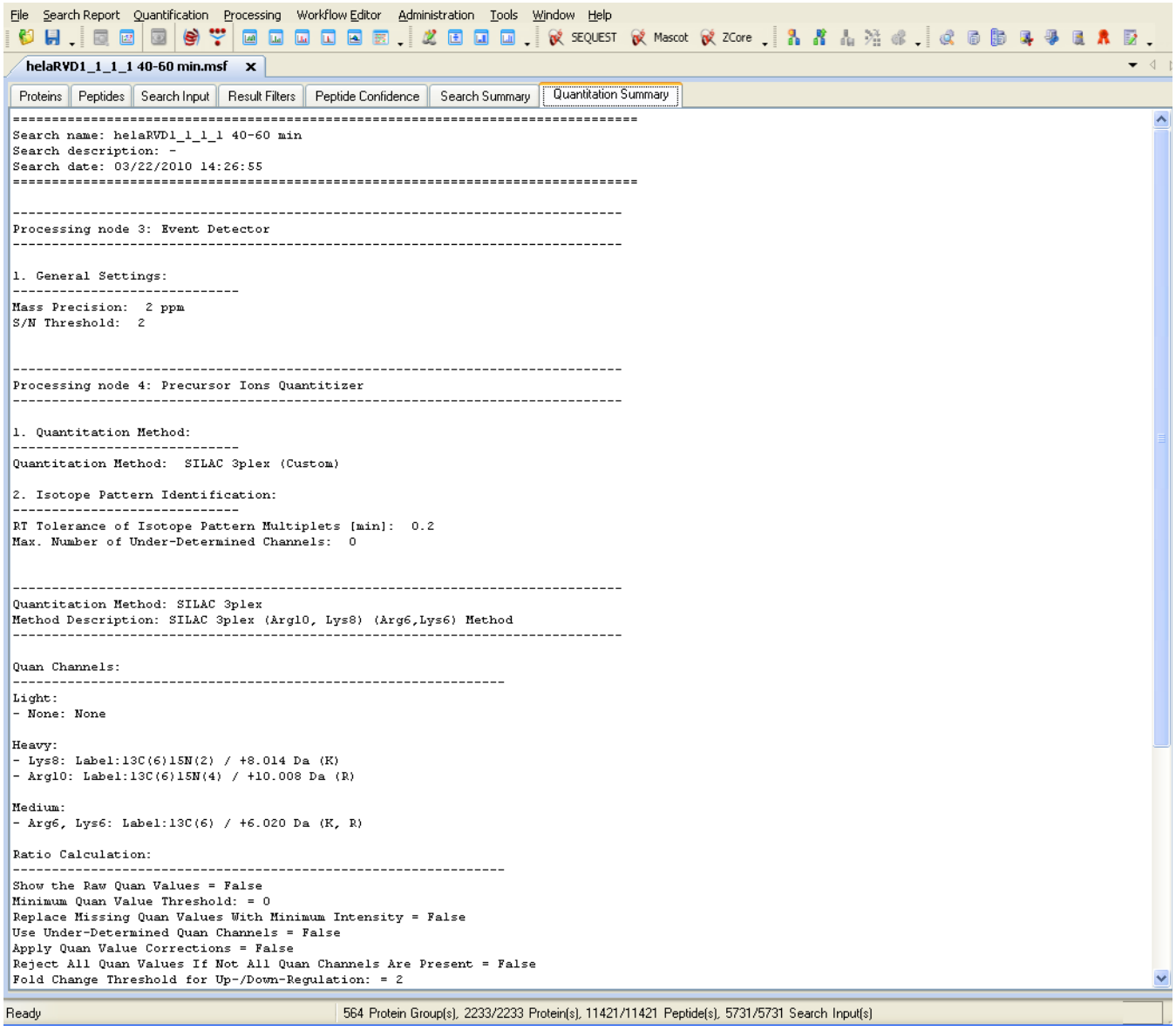
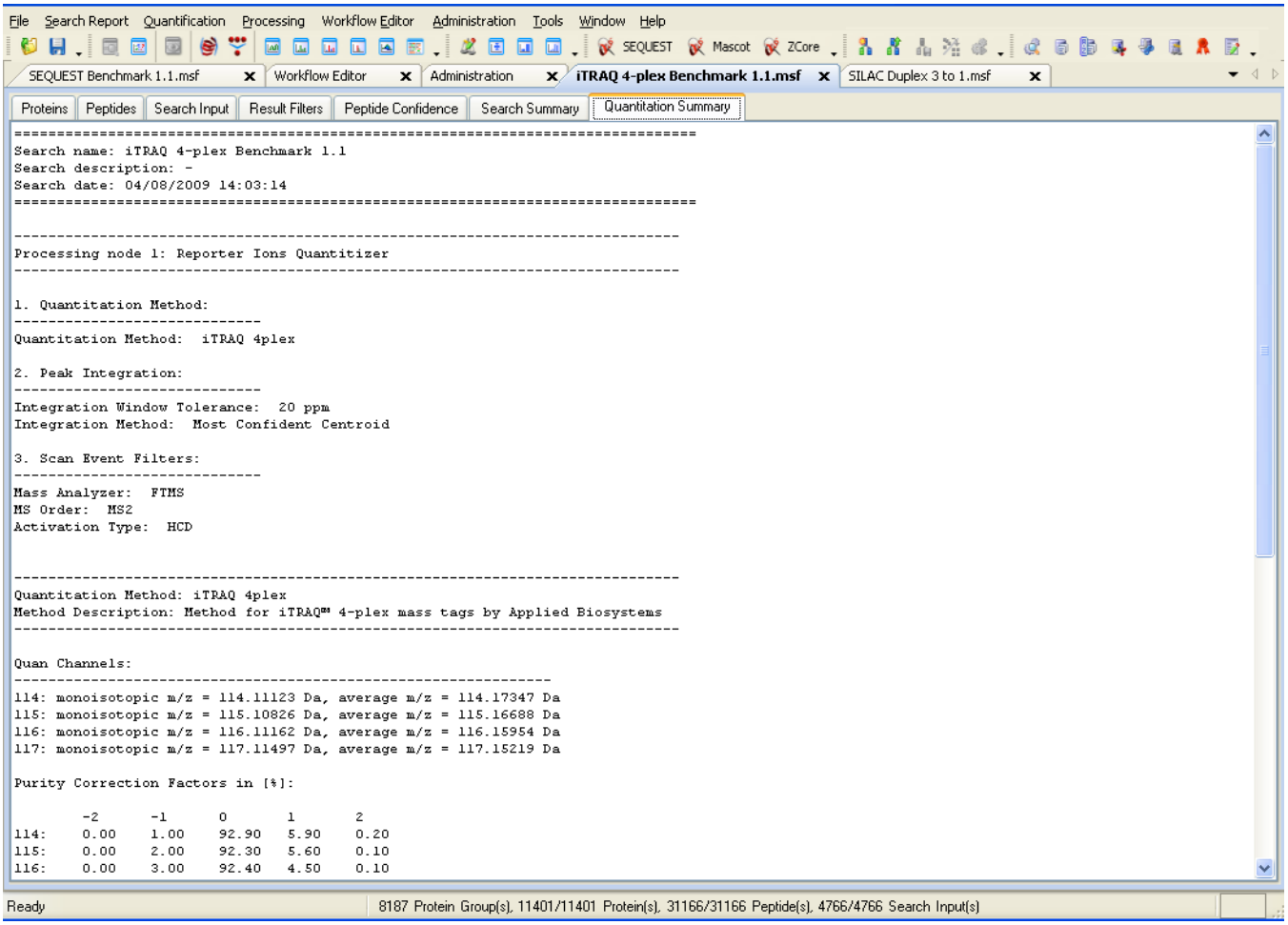


Figure 272. Quantification Summary page for reporter ion quantification



❖ To display the Quantification Summary page

- In an open MSF report, click the **Quantification Summary** tab.

The Quantification Summary page of your search report appears.

Quantification Summary Page Parameters

Table 88 lists the parameters on the Quantification Summary page for precursor ion quantification.

Table 88. Quantification Summary page parameters for precursor ion quantification

Parameter	Description
Processing node section	Lists the quantification nodes used in the workflow and their parameter settings.
Quantification Method section	Displays the quantification method specified by the Quantification Method option in the Parameters pane for the Precursor Ions Quantifier node. For information on this parameter, see "Precursor Ions Quantifier Node" on page 699.
Isotope Pattern Identification section	Displays the settings of the Isotope Pattern Identification options in the Parameters pane for the Precursor Ions Quantification node. For information on these parameters, see "Precursor Ions Quantifier Node" on page 699.
Quantification Method section (Quan Channels, Ratio Calculation, Protein Quantification, Experimental Bias, Ratio Reporting)	Summarizes the settings that you chose on the pages of the Quantification Method Editor dialog box. For information about using this dialog box, see "Setting Up the Quantification Method" on page 485.

Table 89 lists the parameters on the Quantification Summary page for reporter ion quantification.

Table 89. Quantification Summary page parameters for reporter ion quantification (Sheet 1 of 2)

Parameter	Description
Quantification Method section	Displays the quantification method specified by the Quantification Method option in the Parameters pane for the Reporter Ions Quantifier node. For more information on this parameter, see "Reporter Ions Quantifier Node" on page 699.
Peak Integration section	Displays the settings of the Peak Integration options in the Parameters pane for the Reporter Ions Quantifier node. For more information on these parameters, see "Reporter Ions Quantifier Node" on page 699.

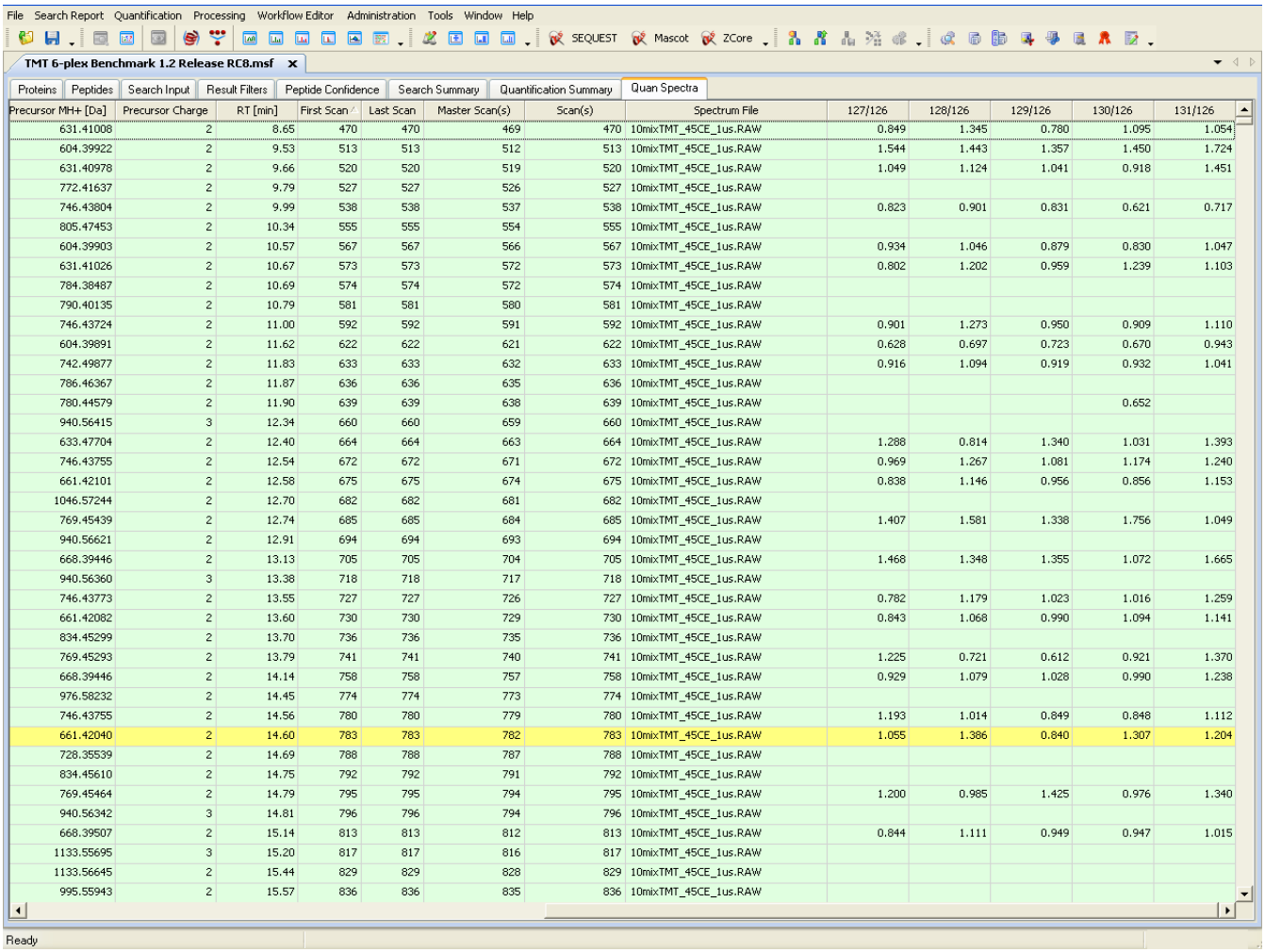
Table 89. Quantification Summary page parameters for reporter ion quantification (Sheet 2 of 2)

Parameter	Description
Scan Event Filters section	Displays the settings of the Scan Even Filters options in the Parameters pane for the Reporter Ions Quantifier node. For more information on these parameters, see "Reporter Ions Quantifier Node" on page 699.
Quantification Method section (Quan Channels, Ratio Reporting, Ratio Calculation, Protein Quantification, Experimental Bias)	Summarizes the settings that you chose on the pages of the Quantification Method Editor dialog box.

Displaying Quantification Spectra

After you perform reporter ion quantification and then select the Show Quan Spectra on Separate Tab check box on the Result Filters page as you open the resulting MSF file, the MSF file includes the Quan Spectra page, shown in Figure 273. This page displays the TMT intensities and ratios for all spectra in reporter ion quantification, regardless of whether they have been identified. This option generates the Quan Spectra page in the MSF report only if you included a Reporter Ion Quantification node in your workflow.

Figure 273. Quan Spectra page



Precursor MH+ [Da]	Precursor Charge	RT [min]	First Scan	Last Scan	Master Scan(s)	Scan(s)	Spectrum File	127/126	128/126	129/126	130/126	131/126
631.41008	2	8.65	470	470	469	470	10mixTMT_45CE_1us.RAW	0.849	1.345	0.780	1.095	1.054
604.39922	2	9.53	513	513	512	513	10mixTMT_45CE_1us.RAW	1.544	1.443	1.357	1.450	1.724
631.40978	2	9.66	520	520	519	520	10mixTMT_45CE_1us.RAW	1.049	1.124	1.041	0.918	1.451
772.41637	2	9.79	527	527	526	527	10mixTMT_45CE_1us.RAW					
746.43804	2	9.99	538	538	537	538	10mixTMT_45CE_1us.RAW	0.823	0.901	0.831	0.621	0.717
805.47453	2	10.34	555	555	554	555	10mixTMT_45CE_1us.RAW					
604.39903	2	10.57	567	567	566	567	10mixTMT_45CE_1us.RAW	0.934	1.046	0.879	0.830	1.047
631.41026	2	10.67	573	573	572	573	10mixTMT_45CE_1us.RAW	0.802	1.202	0.959	1.239	1.103
784.38487	2	10.69	574	574	572	574	10mixTMT_45CE_1us.RAW					
790.40135	2	10.79	581	581	580	581	10mixTMT_45CE_1us.RAW					
746.43724	2	11.00	592	592	591	592	10mixTMT_45CE_1us.RAW	0.901	1.273	0.950	0.909	1.110
604.39891	2	11.62	622	622	621	622	10mixTMT_45CE_1us.RAW	0.628	0.697	0.723	0.670	0.943
742.49877	2	11.83	633	633	632	633	10mixTMT_45CE_1us.RAW	0.916	1.094	0.919	0.932	1.041
786.46367	2	11.87	636	636	635	636	10mixTMT_45CE_1us.RAW					
780.44579	2	11.90	639	639	638	639	10mixTMT_45CE_1us.RAW				0.652	
940.56415	3	12.34	660	660	659	660	10mixTMT_45CE_1us.RAW					
633.47704	2	12.40	664	664	663	664	10mixTMT_45CE_1us.RAW	1.288	0.814	1.340	1.031	1.393
746.43755	2	12.54	672	672	671	672	10mixTMT_45CE_1us.RAW	0.969	1.267	1.081	1.174	1.240
661.42101	2	12.58	675	675	674	675	10mixTMT_45CE_1us.RAW	0.838	1.146	0.956	0.856	1.153
1046.57244	2	12.70	682	682	681	682	10mixTMT_45CE_1us.RAW					
769.45439	2	12.74	685	685	684	685	10mixTMT_45CE_1us.RAW	1.407	1.581	1.338	1.756	1.049
940.56621	2	12.91	694	694	693	694	10mixTMT_45CE_1us.RAW					
668.39446	2	13.13	705	705	704	705	10mixTMT_45CE_1us.RAW	1.468	1.348	1.355	1.072	1.665
940.56360	3	13.38	718	718	717	718	10mixTMT_45CE_1us.RAW					
746.43773	2	13.55	727	727	726	727	10mixTMT_45CE_1us.RAW	0.782	1.179	1.023	1.016	1.259
661.42082	2	13.60	730	730	729	730	10mixTMT_45CE_1us.RAW	0.843	1.068	0.990	1.094	1.141
834.45299	2	13.70	736	736	735	736	10mixTMT_45CE_1us.RAW					
769.45293	2	13.79	741	741	740	741	10mixTMT_45CE_1us.RAW	1.225	0.721	0.612	0.921	1.370
668.39446	2	14.14	758	758	757	758	10mixTMT_45CE_1us.RAW	0.929	1.079	1.028	0.990	1.238
976.58232	2	14.45	774	774	773	774	10mixTMT_45CE_1us.RAW					
746.43755	2	14.56	780	780	779	780	10mixTMT_45CE_1us.RAW	1.193	1.014	0.849	0.848	1.112
661.42040	2	14.60	783	783	782	783	10mixTMT_45CE_1us.RAW	1.055	1.386	0.840	1.307	1.204
728.35539	2	14.69	788	788	787	788	10mixTMT_45CE_1us.RAW					
834.45610	2	14.75	792	792	791	792	10mixTMT_45CE_1us.RAW					
769.45464	2	14.79	795	795	794	795	10mixTMT_45CE_1us.RAW	1.200	0.985	1.425	0.976	1.340
940.56342	3	14.81	796	796	794	796	10mixTMT_45CE_1us.RAW					
668.39507	2	15.14	813	813	812	813	10mixTMT_45CE_1us.RAW	0.844	1.111	0.949	0.947	1.015
1133.55695	3	15.20	817	817	816	817	10mixTMT_45CE_1us.RAW					
1133.56645	2	15.44	829	829	828	829	10mixTMT_45CE_1us.RAW					
995.55943	2	15.57	836	836	835	836	10mixTMT_45CE_1us.RAW					

Quan Spectra Page Parameters

The parameters on the Quan Spectra page are the same as those on the Search Input page (see "Search Input Page Parameters" on page 347), but they also include reporter ion quantification ratio columns that display the corrected ratio of the intensity of the fragmented tag in a sample to the intensity of the fragmented tag in the control sample for all spectra, regardless of whether they have been identified.

The Proteome Discoverer application generates the Quan Spectra page only if you included a Reporter Ion Quantification node in your workflow and select the Show Quan Spectra on Separate Tab check box on the Result Filters page when you open an MSF file. For more information on generating this page, see "To filter the data in the MSF file before opening it" on page 218.

Displaying Quantification Channel Values

You can generate a chart that displays the absolute intensity (for reporter ion quantification) or the area (for precursor ion quantification) of the quantification values detected for the available quantification channels.

❖ **To display quantification channel values**

1. Click the row of the peptide that interests you.

   To obtain meaningful results, "Used" must appear in the Quan Info column of the report.

2. Choose **Quantification > Show Quan Channel Values**, or click the **Show Quan Channel Values** icon,  .

Displaying Quantification Channel Values for Reporter Ion Quantification

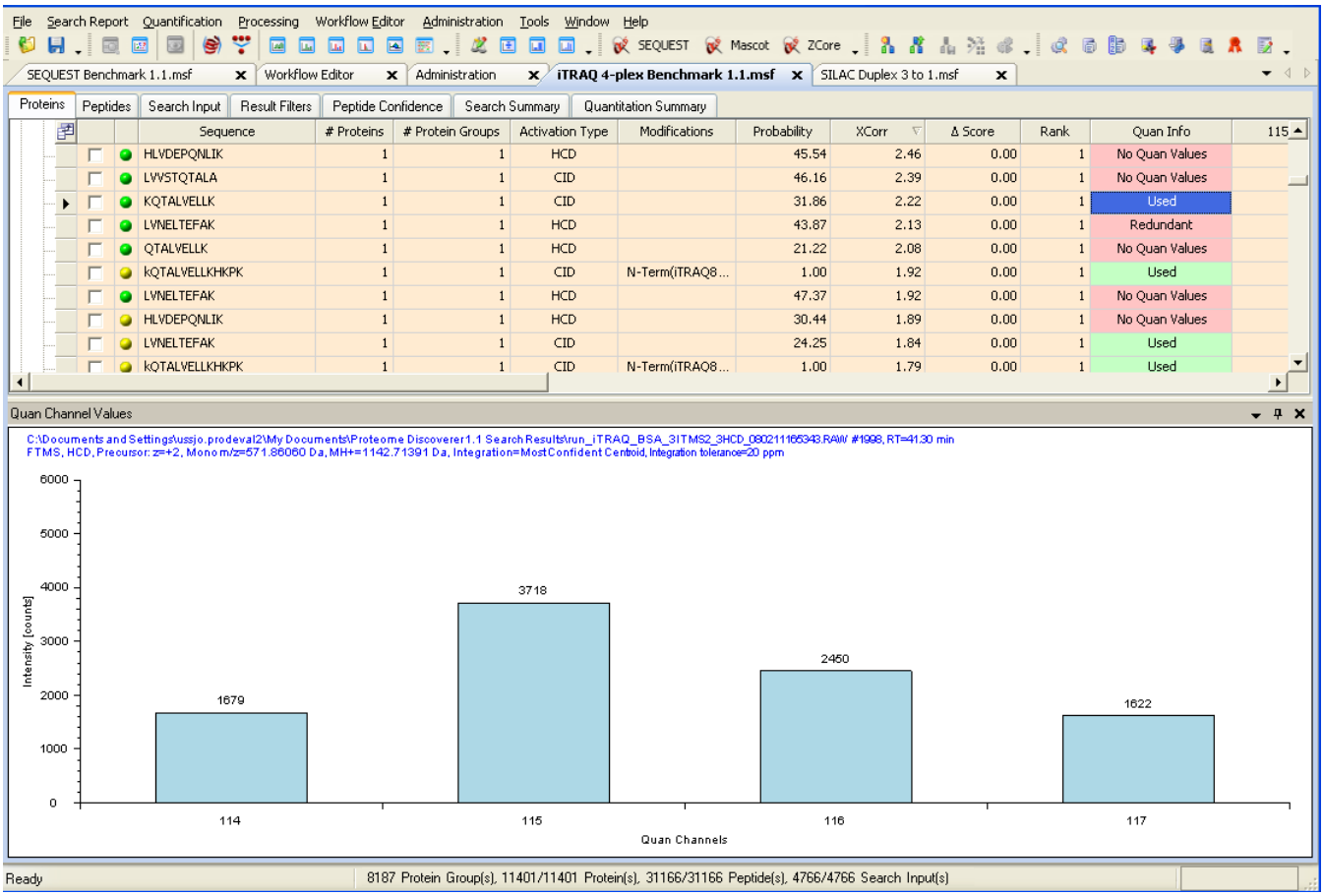
For reporter ion quantification, you can generate a Quantification Channel Values chart that displays the absolute intensity of the reporter ions detected for the available quantification channels.

Reporter ions, or reporters, are the labels affixed to peptide samples in reporter ion quantification. They fragment in the MS/MS process. You can use the quantification value intensity to calculate the relative ratio of a peptide. You might also want to view the absolute quantification value intensity to verify that the peptide ratio calculation is correct.

The *x* axis of the chart shows the names of the quantification channels, and the *y* axis shows the intensity of the reporter ions, in counts.

The 4plex quantification method in iTRAQ has four reporter ions. Suppose that they are used to label four biological samples: 114, 115, 116, and 117. Figure 274 shows the quantification Channel Values chart created by the Show Quan Channel Values command for these samples. It shows the relative intensities of the samples labeled with the 114, 115, 116, and 117 reporter ions. Clearly, the sample labeled 115 is the sample with the greatest reporter ion intensity.

Figure 274. Quantification Channel Values chart for reporter ion quantification



Displaying Quantification Channel Values for Precursor Ion Quantification

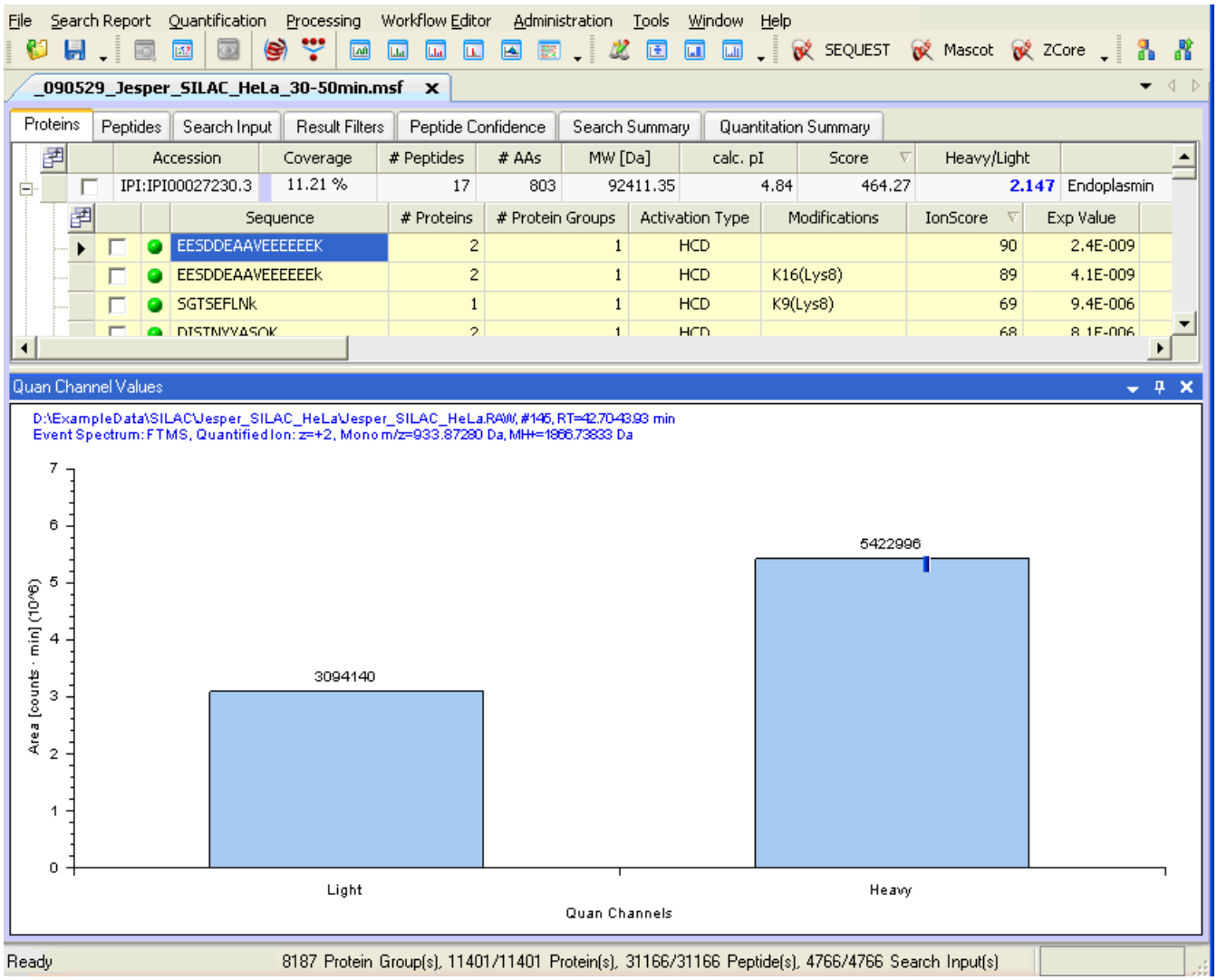
For precursor ion quantification, you can generate a Quantification Channel Values chart that displays the area of the isotopes detected for the available quantification channels.

Heavy isotopes are incorporated into proteins in precursor ion quantification. You can use the quantification value area to calculate the relative ratio of a peptide. You might also want to view the quantification value area to verify that the peptide ratio calculation is correct.

The *x* axis of the chart shows the quantification channels, and the *y* axis shows the detected area for the given quantification channel, defined by counts per minute.

The Lysine 8/Arginine 10 quantification method in SILAC has a sample labeled with the Light isotope and a sample labeled with the Heavy isotope. Figure 275 shows the chart created by the Show Quan Channel Values command for these samples. It shows the relative area of the samples labeled with the Light and Heavy isotopes. Clearly, the sample labeled Heavy is the sample with the greatest area.

Figure 275. Quantification Channel Values chart for precursor ion quantification



Quantification Channel Values Chart Parameters

Table 90 describes the features of the Quantification Channel Values chart.

Table 90. Quantification Channel Values chart parameters

Parameter	Description
Header	Displays the following: • Name and location of the raw file • Numbers of any grouped spectra • Retention time of the precursor ion • Mass spectrometer used • Activation type used • Mass-to-charge ratio (*m/z*) of the "light" ion (precursor ion quantification) or the mass-to-charge ratio of the precursor ion (reporter ion quantification), in daltons • MH+ (the protonated monoisotopic mass), in daltons • Integration method (reporter ion quantification) • Integration tolerance (reporter ion quantification)
Quan Channels	Specifies the names of the quantification channels used.
Area [counts × min]	Specifies the area of the extracted ion chromatogram of each channel. The relative peak intensities of multiple isotopically distinct peptides from each protein are then used to determine the average change in protein abundance in the treated sample.

Displaying the Quantification Spectrum Chart

You can generate a chart showing the spectrum used for quantification. This chart is available for every peptide with an associated quantification result.

❖ **To display the Quantification Spectrum chart**

1. Select the peptide of interest.

   The peptide must be labeled "Used" in the Quan Usage column of the Peptides page.

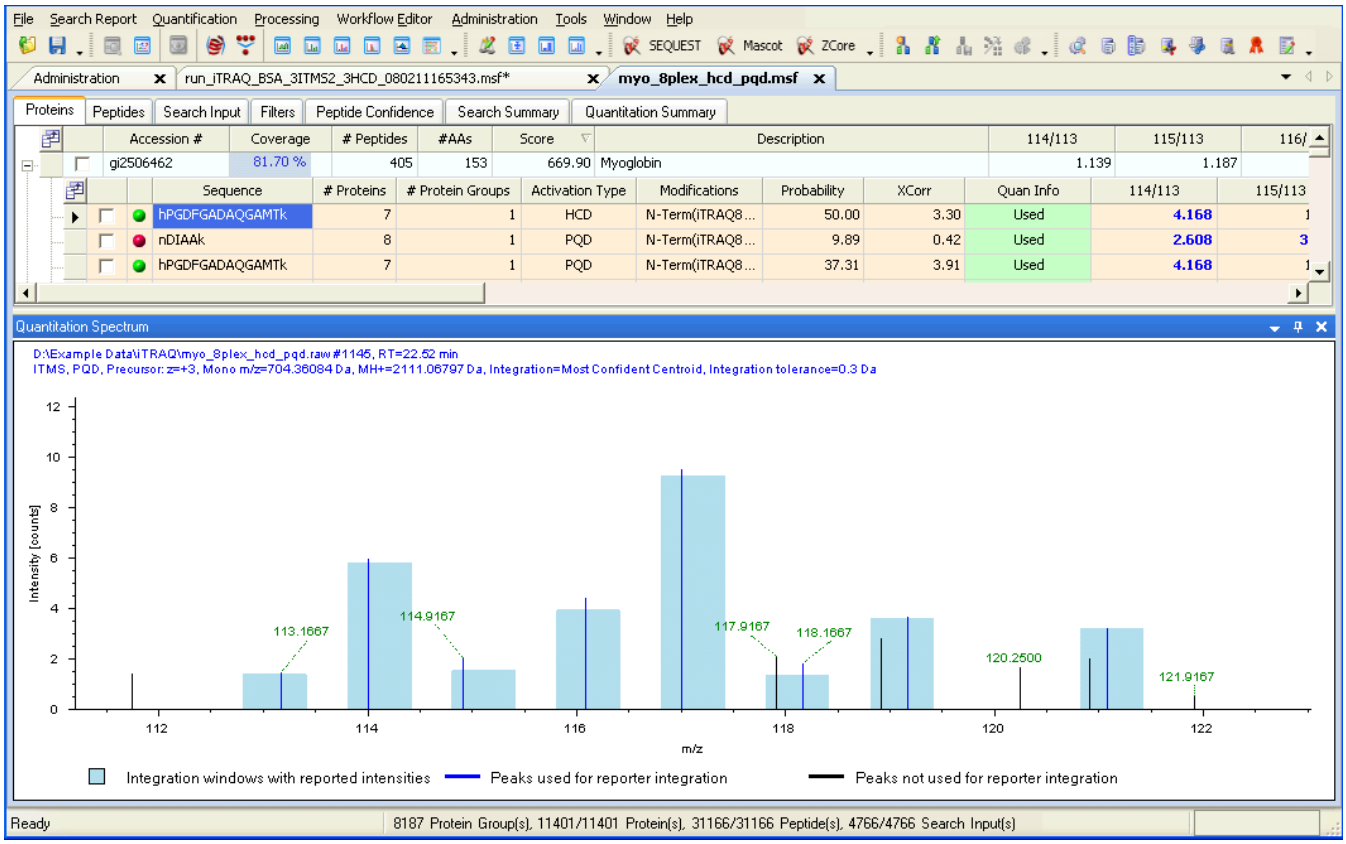
2. Choose **Quantification > Show Quantification Spectrum**, or click the **Show Quantification Spectrum** icon,  .

Displaying the Quantification Spectrum Chart for Reporter Ion Quantification

For reporter ion quantification, the Quantification Spectrum chart displays the intensity of the reporter ions, in counts. It shows a spectrum for each peptide, except for those peptides labeled with "No Quan Values."

Figure 276 shows an example of a quantification spectrum from an iTRAQ 8plex sample quantified with an Integration Window Tolerance setting (in the Reporter Ions Quantifier node) of 0.3 Da for extracting the reporter peaks from the quantification spectrum.

Figure 276. Quantification Spectrum chart for an iTRAQ 8plex sample using Proteome Discoverer scans



The Quantification Spectrum chart includes the following features:

- The light blue boxes represent the integration windows for the reporter tags. The boxes are centered on the masses of the reporter tags, as specified in the quantification method. The width of boxes is the integration window used for extracting the reporter tags. It is ±0.3 Da, as specified by the settings of the parameters in the Reporter Ions Quantifier node (you can look up all these values on the Quantification Summary page). The height of the line in the box represents the actual tag intensity used for calculating the peptide ratios. The height of the box represents the corrected tag intensity. The height depends on the setting of the Integration Method parameter specified in the Reporter Ions Quantifier node. It is always the value that results from correction for isotopic impurities, as specified in the Reporter Ion Isotopic Distribution area of the Quan Channels page of the Quantification Method Editor dialog box, shown in Figure 257 on page 493.
- To calculate the actual intensity of a particular tag, the Proteome Discoverer application chooses the blue fragment peaks from the spectrum, and considers only peaks in the integration window.

- The black fragment peaks represent peaks that are present in the spectrum but that are not chosen for calculating the tag intensities. They might not be chosen because the peaks lie outside of any integration window, or because the setting of the Integration Method parameter specified in the Reporter Ions Quantifier node determined that only one peak per integration window should be chosen from any integration window. A different peak was picked for this integration window according to the criterion specified by the Integration Method setting.

Missing Reporter Peaks in the Quantification Spectrum

If reporter ions are missing in the quantification spectra, you can use settings on the Ratio Calculation page of the Quantification Method Editor dialog box to influence how the Proteome Discoverer application handles this problem. For example, if all four intensities of an iTRAQ 4plex are missing, or if the reference ion is missing (for example, in the iTRAQ 4plex method shown in Figure 277, the 114 ion is missing), the corresponding spectrum is always excluded from the protein quantification. In the Quan Info column of the Peptides page, these peptides are marked “No Quan Values” as shown in Figure 277. This figure shows the quantification results for a protein where two of its spectra showed no reporter ions at all, one peptide was redundant, and two additional spectra had missing reporter ions. The protein ratios were calculated according to the settings displayed in Figure 278.

Figure 277. Quantification results with missing reporter ions

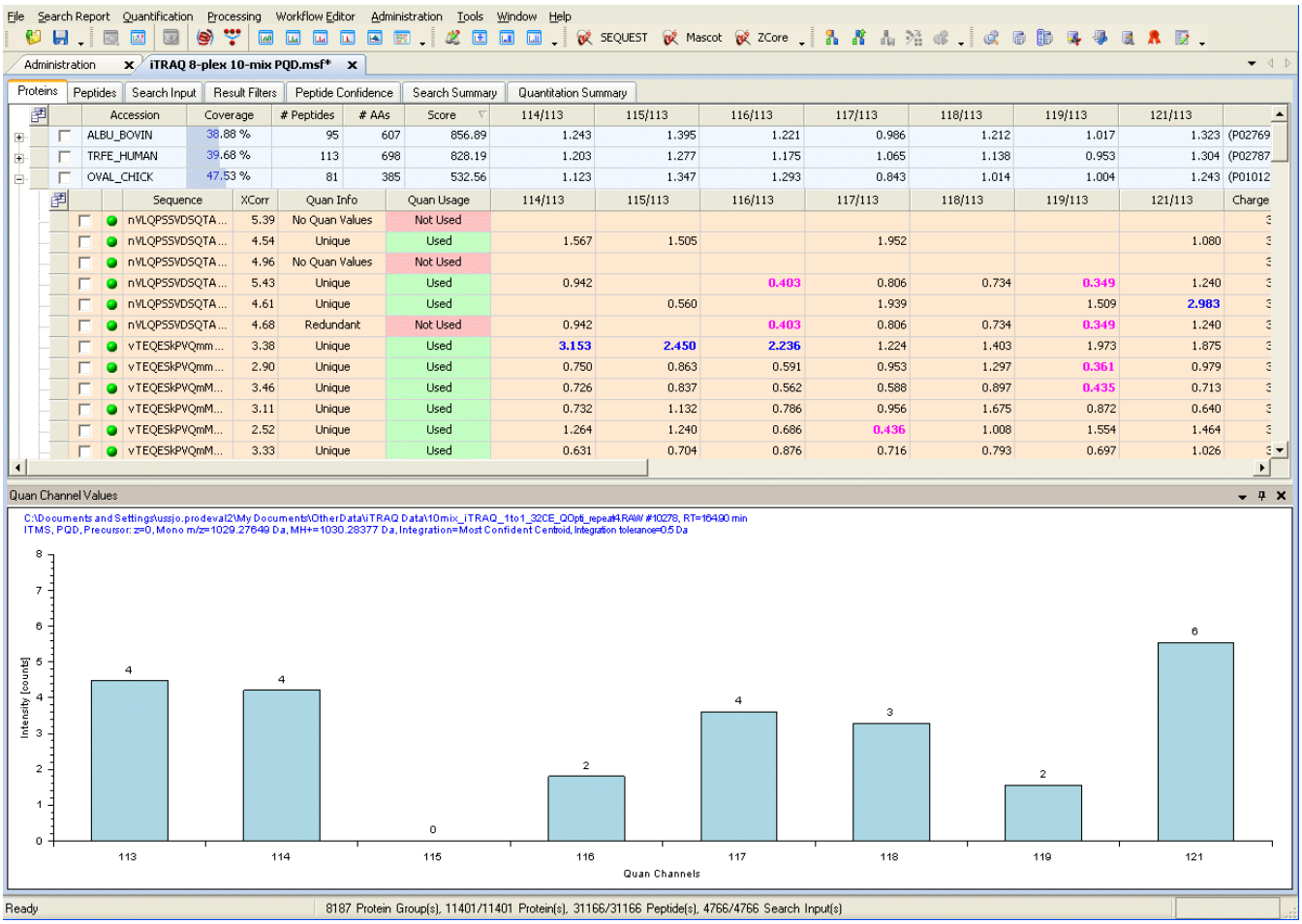
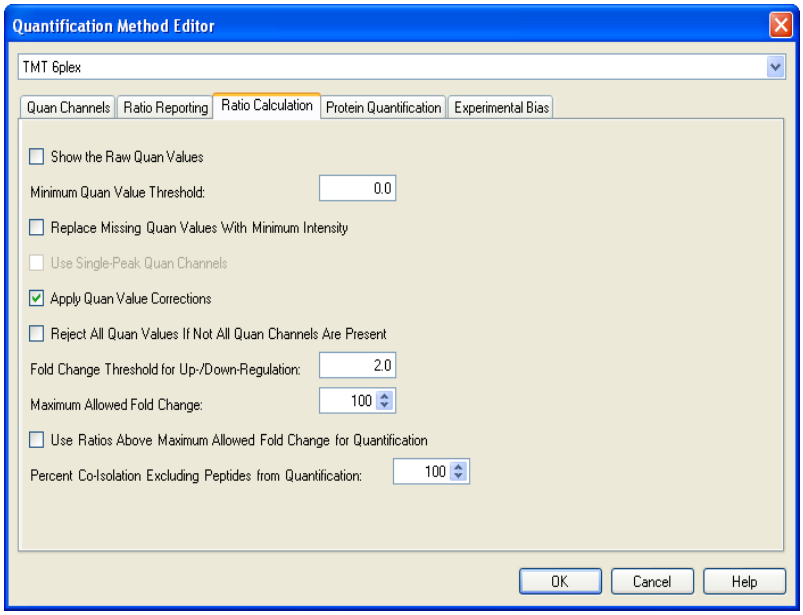


Figure 278. Ratio Calculation page of the Quantification Method Editor dialog box



For information about the options on the Ratio Calculation page of the Quantification Method Editor dialog box, see "Setting Up the Ratio Calculation" on page 497 or "Ratio Calculation Page Parameters" on page 511.

If one or more of the reporter, or mass, tags are missing in the quantification spectrum, the calculated ratios are either zero or infinity, depending on which tag intensity is the numerator and which is the denominator. Even if all tags are present, the calculated ratios might be very high or very low. You can use the Maximum Allowed Fold Change option on the Ratio Calculation page of the Quantification Method Editor dialog box to replace such extremely high or extremely low ratios with the maximum allowed number of times that the ratios can be multiplied. In the example in Figure 278, the maximum allowed number of times that the ratio can be multiplied is 100. That is, extremely high ratios are replaced by 100, and extremely low ratios are replaced by 0.01. There is an inherent dynamic range that you can expect to be valid or detectable with the given instrumentation and method.

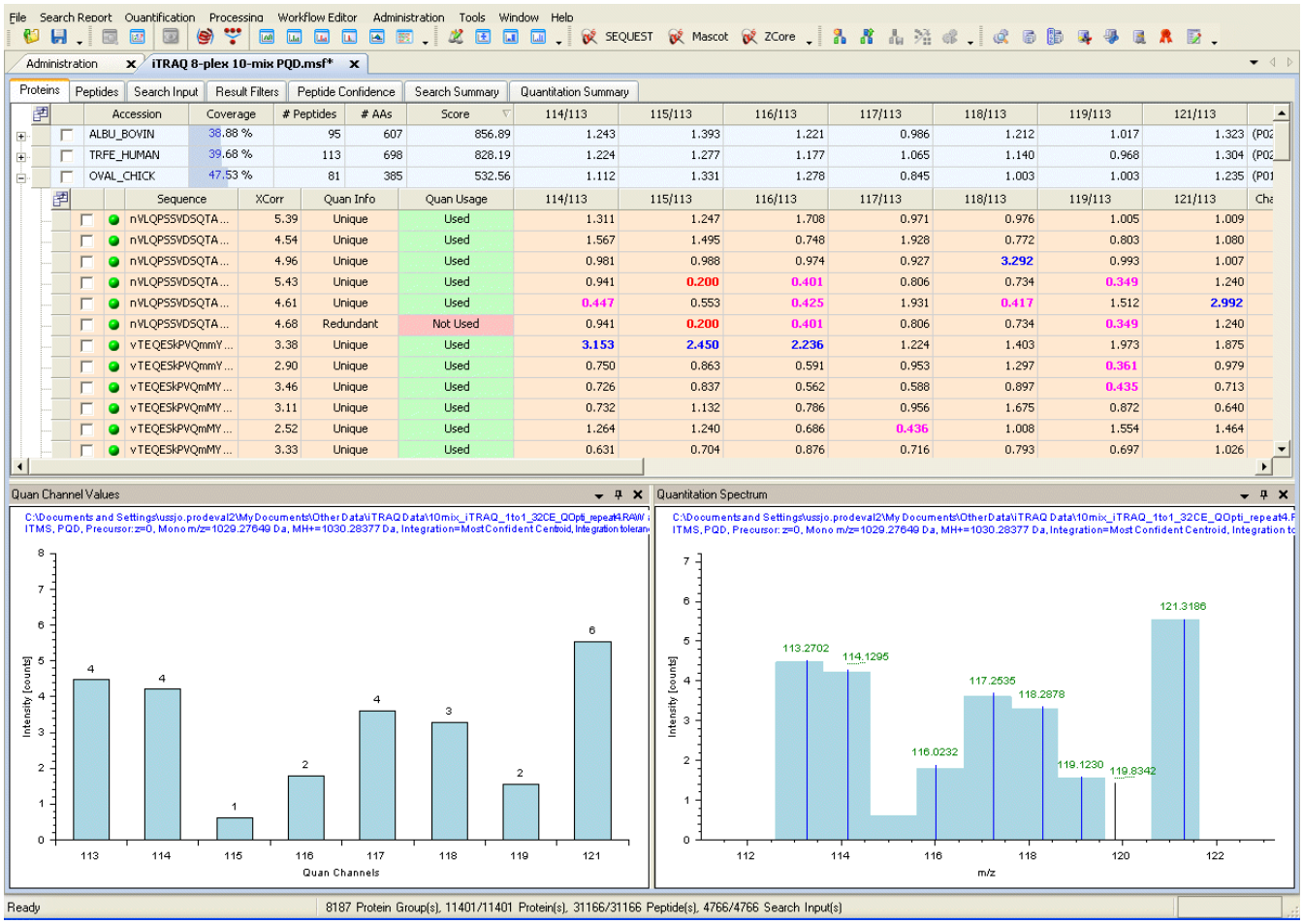
The Use Ratios Above Maximum Allowed Fold Change for Quantification option in the Ratio Calculation dialog box specifies whether such maximum calculated ratios should be considered when the Proteome Discoverer application calculates the protein ratios. You can use this option to automatically include extreme values when the application calculates the protein ratios. Since the protein ratios are calculated as the median, outlier protein ratios are likely to occur only if there is a sufficient number of peptides that can be used for protein quantification.

In the example in Figure 277, the protein is identified with 81 peptides. For two of the peptides, the corresponding quantification spectra show no reporter tag at all. These peptides are never considered in calculating the protein ratios and are marked as "No Quan Values" in the Quan Info column of the Peptides page. Two additional spectra are missing individual reporter ions. Although two peptides are marked as "Used" for quantification in the Quan Info column, their extreme ratios are not considered in the protein ratio calculation (with the settings in Figure 278).

If at least one of the reporter intensities is present (see Figure 277), you can use the Replace Missing Reporter Intensities With Minimum Intensity option on the Ratio Calculation page of the Quantification Method Editor dialog box to replace the missing intensities with the minimum intensity detected among all spectra on all reporter channels. Reporter intensities are missing because they fall under the detection limit, so it might make sense to replace them with an intensity estimate that is close to the detection limit.

Figure 279 shows the same protein as in Figure 277 after the selection of the Replace Missing Reporter Intensities with Minimum Intensity option. In the example, the 115 reporter ion has been replaced with a minimum intensity value. This is not exactly the true value, but it is better than having no estimates for the ratios of this protein. Whether this option gives valuable results for you depends on your experimental design and quantification strategy.

Figure 279. Quantification results after applying the Replace Missing Reporter Intensities with the Minimum Intensity option

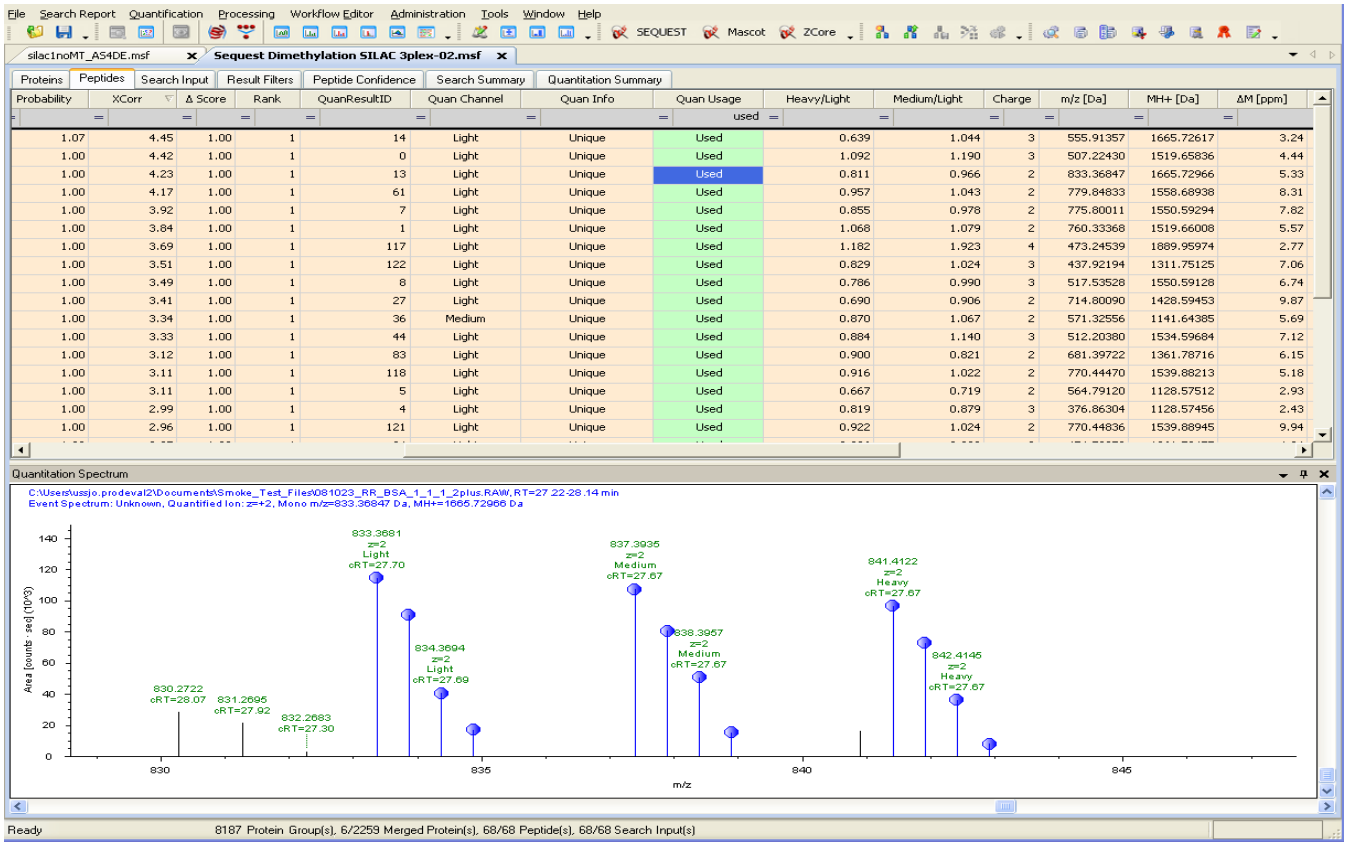


You exclude spectra with one or more missing reporter peaks from the protein ratio calculation by selecting the Reject All Quan Values If Not All Quan Channels Are Present option on the Ratio Calculation page.

Displaying the Quantification Spectrum Chart for Precursor Ion Quantification

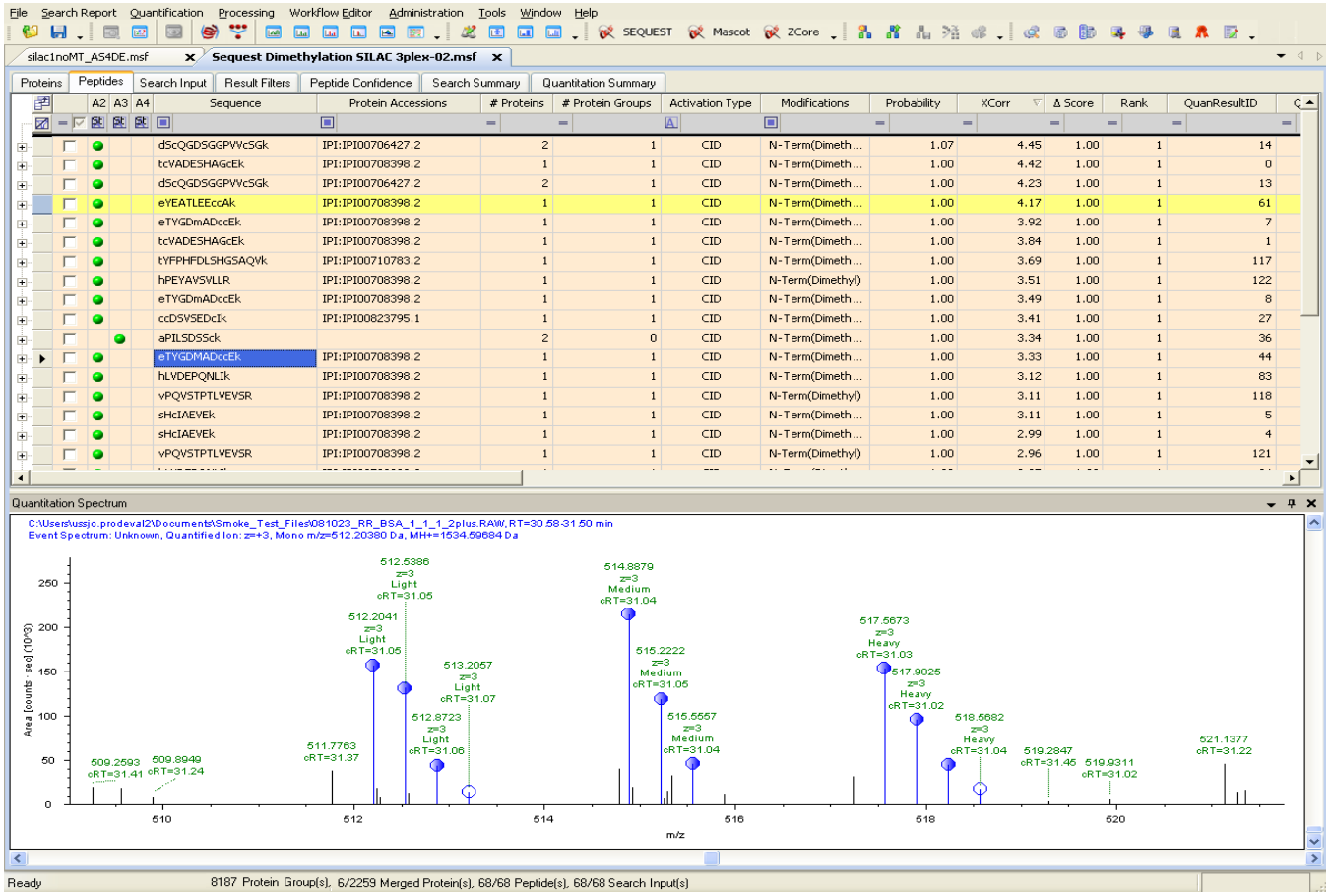
For precursor ion quantification, the Quantification Spectrum chart displays a quantification spectrum for each peptide. It also displays the different abundances of the identified Light, Medium, and Heavy isotopic peak patterns used to quantify a peptide. The abundances are measured by calculating the area of the extracted ion chromatogram of each isotope of a pattern. The chart highlights the corresponding isotope pattern peaks and labels them with the quantification channel names, as shown in Figure 280. It also includes any peaks that are not part of an isotope pattern.

Figure 280. Quantification Spectrum chart for precursor ion quantification



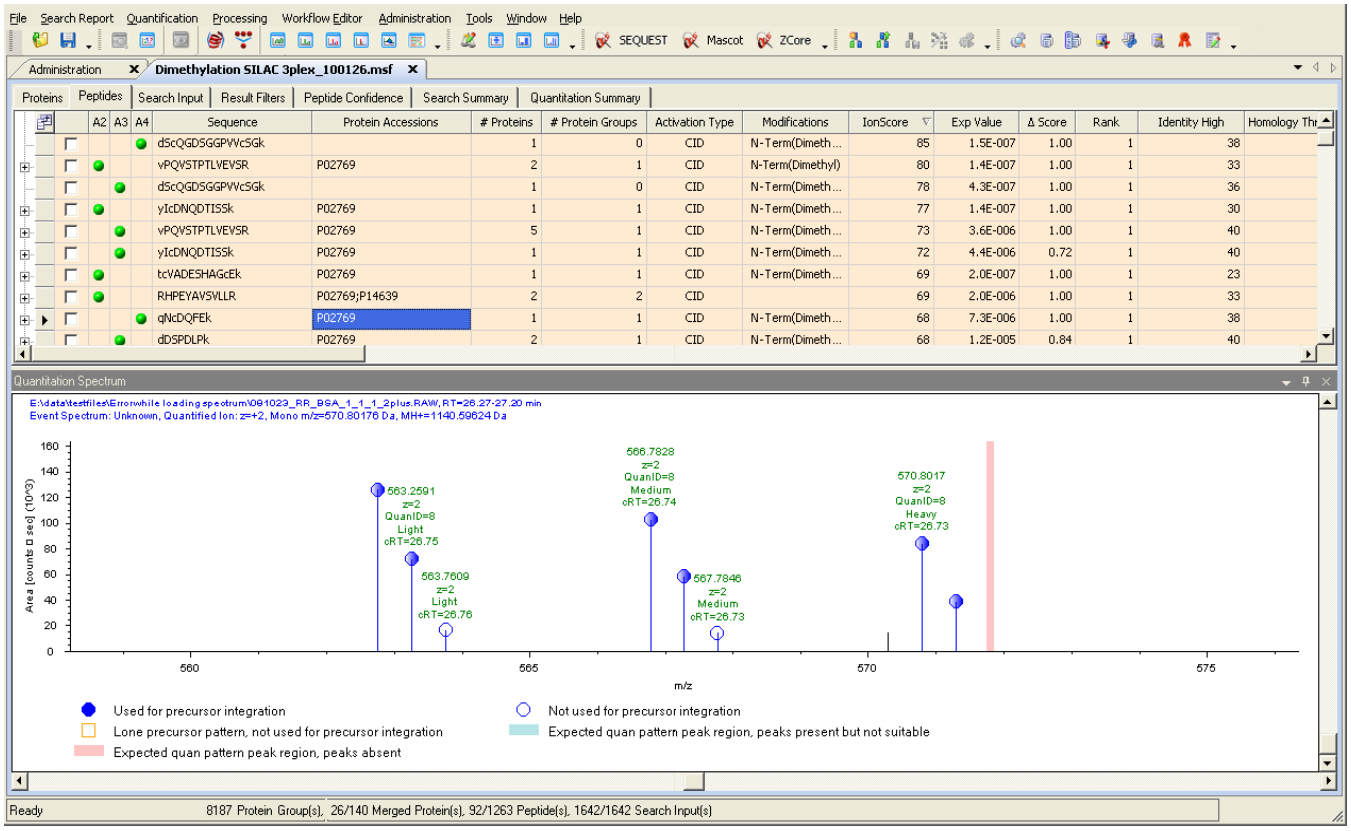
The *x* axis of the chart displays the mass-to-charge ratio of the isotopes, and the *y* axis displays the area of the extracted ion chromatogram for the isotopes. Filled blue circles mark the isotope pattern peaks that were used for calculating the quantification values for the different quantification channels. Unfilled blue circles mark the isotope pattern peaks that were identified but not used. The Quantification Spectrum chart always compares the exact same isotopic pattern peaks for each label. For example, the chart in Figure 281 compares the first three isotopic pattern peaks among all three types: Light, Medium, and Heavy. But the chart also contains an additional Light isotopic pattern peak and an additional Heavy isotopic pattern peak that are not used, so these two peaks are represented by unfilled circles.

Figure 281. Extra isotopic pattern peaks represented by unfilled circles in the Quantification Spectrum chart



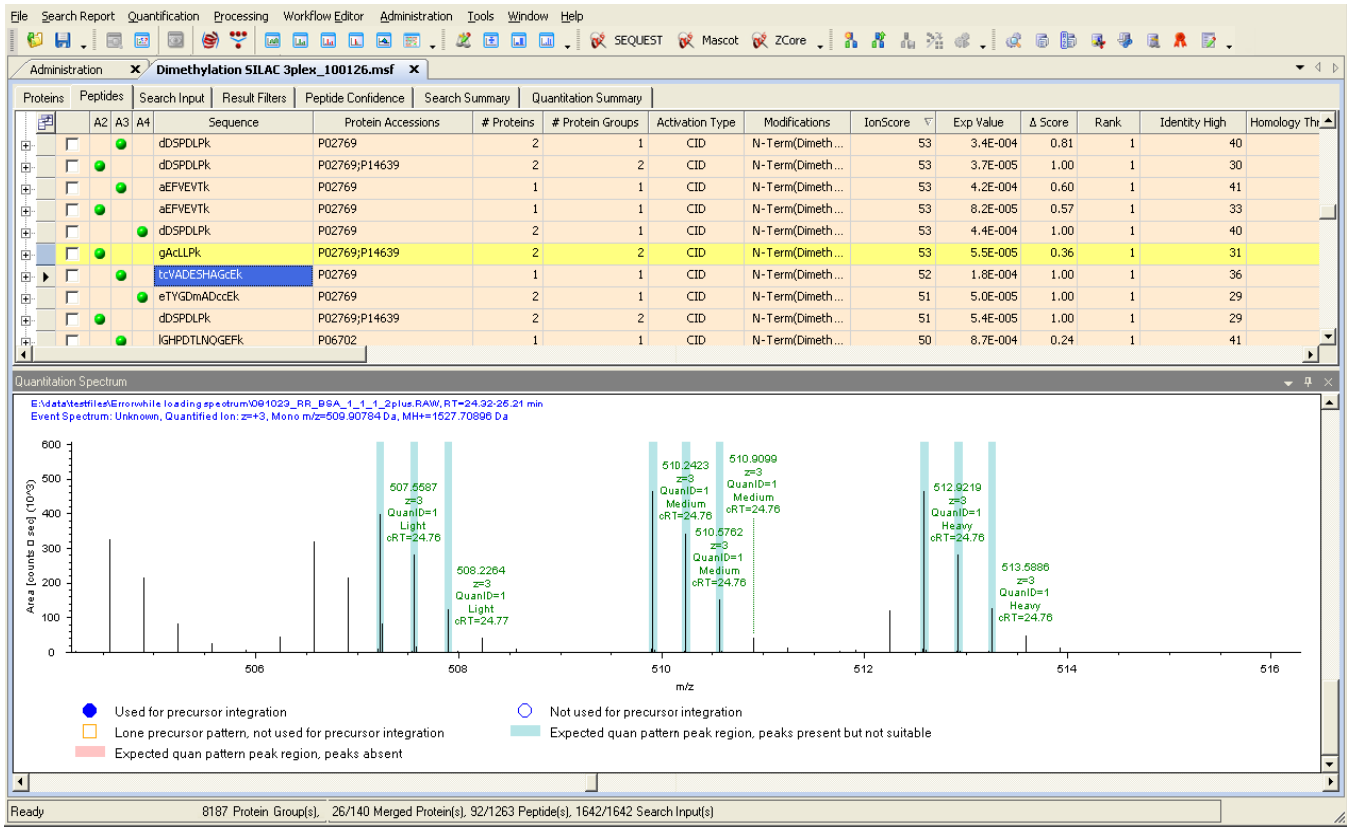
The Quantification spectrum chart can also indicate whether an expected quantification pattern peak is absent. Regions in pink indicate where a quantification pattern peak was expected but is absent (see Figure 282). This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Figure 282. Expected but absent peak in the Quantification Spectrum chart



Regions in blue, shown in Figure 283, indicate where a quantification pattern peak was expected but is unsuitable. Pattern peaks might be unsuitable because of the wrong centroid retention time, out of delta mass range, wrong intensity, or a peak that has been used by another isotopic pattern. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Figure 283. Expected but unsuitable peaks in the Quantification Spectrum chart



The legends on the Quantification Spectrum chart in Figure 282 and Figure 283 show what the various colors mean. Table 91 also defines the meanings of these colors.

Table 91. The meaning of colors in the Quantification Spectrum chart (Sheet 1 of 2)

Color	Meaning
Filled blue circle	Indicates the isotope pattern peaks that are used in calculating the quantification values for the different quantification channels.
Unfilled blue circle	Indicates the isotope pattern peaks that are not used in calculating the quantification values for the different quantification channels.
Yellow box	Indicates that the pattern includes peaks from only one channel. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Table 91. The meaning of colors in the Quantification Spectrum chart (Sheet 2 of 2)

Color	Meaning
Pink bar	Indicates that a quantification pattern peak is expected but is missing. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.
Blue bar	Indicates that a quantification pattern peak is present but is unsuitable because of errors in peptide labeling or because of the wrong centroid retention time, out of delta mass range, wrong intensity, or a peak that has been used by another isotopic pattern. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Quantification Spectrum Chart Parameters

Table 92 describes the features of the Quantification Spectrum chart.

Table 92. Quantification Spectrum chart parameters

Parameter	Description
Header	Displays the following: • Name and location of the raw file • Scan numbers of any grouped spectra • Retention time of the precursor ion • Type of mass spectrometer used • Activation type (reporter ion quantification only) • Mass-to-charge ratio of the precursor ion, in daltons (reporter ion quantification) or the mass-to-charge ratio of the light isotope of the peptide (precursor ion quantification). • Monoisotopic mass of the peptide of interest • MH+ (the protonated monoisotopic mass), in daltons • Integration method (reporter ion quantification only) • Integration tolerance (reporter ion quantification only)
m/z	Displays the mass-to-charge ratio of the reporter ion tags (for reporter ion quantification) or the isotopes (for precursor ion quantification).
Intensity [counts]/ Area [counts × min]	Displays the intensity of the reporter ions, in counts, (for reporter ion quantification) or the extracted chromatographic area of the isotopes (for precursor ion quantification).

Using Reporter Ion Isotopic Distribution Values to Correct for Impurities

iTRAQ and TMT kits consist of labels that contain different numbers of ^{13}C atoms, ^{15}N atoms, or both. For simplicity, assume that a 4plex kit yields peaks at 114, 115, 116, and 117 *m/z*, which correspond to $^{13}C1$, $^{13}C2$, $^{13}C3$, and $^{13}C4$, respectively. Because the label substances are not 100 percent isotopically pure, each label contains a certain number of other atoms. For example, the 116 label would not consist only of label molecules having three ^{13}C atoms but might also contain label molecules with only one or two ^{13}C atoms or even four or five ^{13}C atoms. As a result, these impurities lead to an observed peak of 116 *m/z*, which is smaller than might be expected if the tag were 100 percent isotopically pure, and to additional peaks at positions –2, –1, +1, +2 Da apart from 116 *m/z*. The intensities of the latter peaks are proportional to the amount of the described isotopic impurities. When the 116 label and the 114, 115, and 117 labels are used, these latter three labels contribute to the peak at 116 *m/z* because of their isotopic impurities.

The intensity of the peak at 116 *m/z* effectively includes the following contributions:

(*observed intensity 116*) = (*true intensity 116*) – (*intensity loss because of 116 impurities*) + (*intensity gain because of other label impurities*)

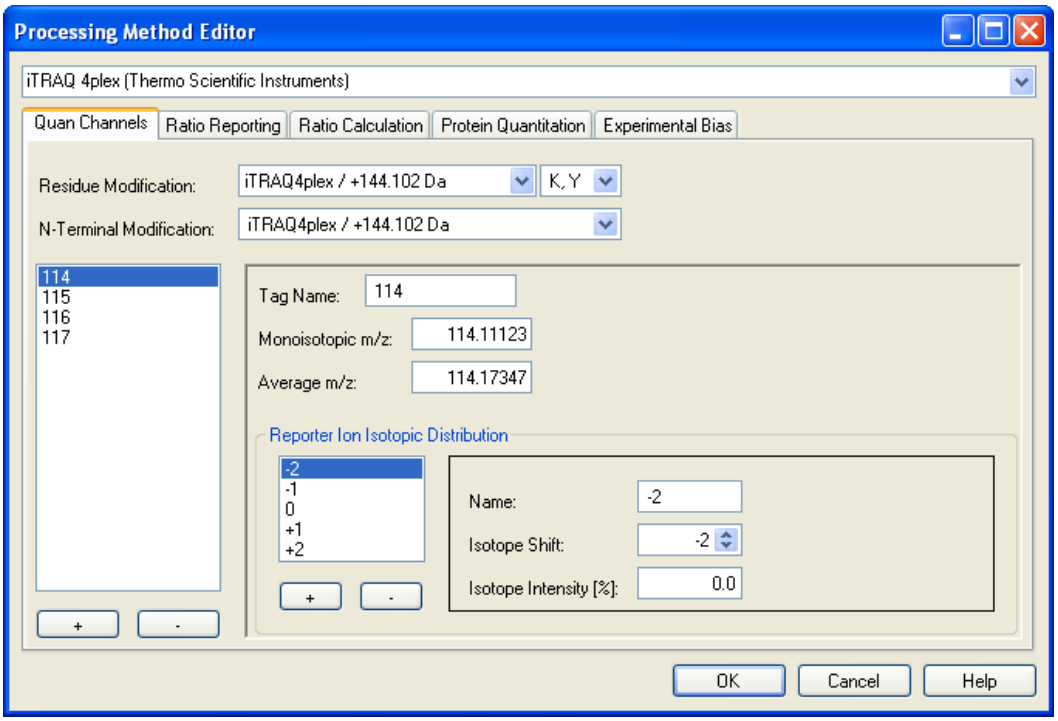
To obtain the true intensity value of the 116 label—that is, the amount of the substance initially labeled with the 116 tag—you must correct the experimentally observed peak for the impurity of the labels.

For a 4plex sample, there are four formulas that use the equation just given for each of the labels, and the proper correction would consider both contributions in the formula by solving the system of coupled linear equations:

(*intensity_of_loss_because_of_116_impurity* and *intensity_of_gain_because_of_other_label_impurities*)

For this correction, you must enter the isotopic distribution of each of the labels used in the quantification method, as shown in Figure 284. The values are part of each of the iTRAQ or TMT label kits used.

Figure 284. Entering values for the isotopic distribution of a specific reporter tag



You can also deconvolute the overlapping labels using other methods. Compatible with the Mascot search engine from Matrix Science, the Proteome Discoverer application uses a first-order approximation to the solution. The error made is small when the intensities of all possible contributing labels are of similar height, and it becomes larger if the intensity differences become larger.

Excluding Peptides from the Protein Quantification Results

You can manually exclude and include certain peptides from the protein quantification results. You can also return excluded or included peptides to their default status.

You cannot include peptides if No Quan Values, Inconsistently Labeled, or Excluded by Method appears in the Quan Info column.

❖ To exclude a peptide from the quantification results

1. On the Proteins page, click the plus sign (+) next to the protein of interest to display its constituent peptides.

2. Right-click the peptide of interest, which must display "Used" in the Quan Usage column, and choose **Include/Exclude Peptide(s) from Protein Quantification > Exclude** from the shortcut menu.

   "Not Used (Excluded)" now appears in the Quan Usage column.

3. To save the information in the MSF file resulting from this setting, choose **File > Save Report**.

❖ To include an excluded peptide in the quantification results

1. On the Proteins page, click the plus sign (+) next to the protein of interest to display its constituent peptides.
2. Right-click the peptide of interest, which must display "Not Used (Excluded)," "Redundant," or "Not Unique" in the Quan Info column, and choose **Include/Exclude Peptide(s) from Protein Quantification > Include** from the shortcut menu.

   "Used (Included)" now appears in the Quan Usage column.

3. To save the information in the MSF file resulting from this setting, choose **File > Save Report**.

❖ To return an included or excluded peptide to its default status

1. On the Proteins page, click the plus sign (+) next to the protein of interest to display its constituent peptides.
2. Right-click the peptide of interest and choose **Include/Exclude Peptide(s) from Protein Quantification > Default** from the shortcut menu.

   The Quan Usage column now displays the peptide's usage status when the MSF file was first opened.

Excluding Peptides with High Levels of Co-Isolation

To create a fragment spectrum, you select a precursor mass for isolation, isolate and fragment the ions within a mass window that you define, and record the product ion masses created.

Ideally, you would isolate and fragment only the precursor ions of a single selected component. However, in practice you isolate the precursor ions within a user-specified window—typically 1 or 2 daltons around the isolation mass. Co-eluting components with a mass falling into this isolation window are also isolated and fragmented. This process is called co-isolation. The co-isolating components are likely to be peptides whose fragments are observed in the created fragment spectra. The co-isolation can exacerbate the identification of the selected peptide and lower the identification confidence.

Co-isolation is an issue in reporter ion quantification. In this type of quantification, the peptides from different charges of the same sample—for example, different treatment states—are modified with special isobaric labels. The isobaric labels disaggregate during precursor ions fragmentation and create reporter tags that appear in the low-mass region of the fragment spectra. The intensity ratio of the observed fragment tags is used for relative quantification of the peptides from the different sample charges.

The co-isolating peptides also create reporter tags that superimpose on the reporter tags of the selected peptide. Because most of the proteins in a real sample are unregulated, the co-isolated peptides often create reporter tags with equal intensity. If these superimpose on the reporter tags of a selected peptide of a regulated protein, the observed ratios of the reporter tags in the fragment spectra can be false. Furthermore, the perturbed ratios of the selected peptides that are greatly affected by co-isolation can also adversely affect the ratios that the Proteome Discoverer application calculates for the proteins that include these peptides.

Determining the extent to which the real reporter tag ratios of the selected peptides are perturbed is difficult. It depends on the level of co-isolation and the isolation characteristics of the instrument. The Proteome Discoverer application flags PSMs with a high level of co-isolation. For newly generated MSF files, it calculates and displays the percentage of interference within the precursor isolation window. This percentage is the relative amount of ion current within the isolation window that is not attributed to the precursor itself:

$$\%_isolation_interference = 100 \times \left[1 - \left(\frac{precursor_intensity_in_isolation_window}{total_intensity_in_isolation_window}\right)\right]$$

The application displays the calculated interference value in the % Isolation Interference column on the Peptides and Search Input pages. For reporter ion quantification, a high isolation interference value could indicate that a calculated peptide ratio is skewed by the presence of co-isolated peptide species.

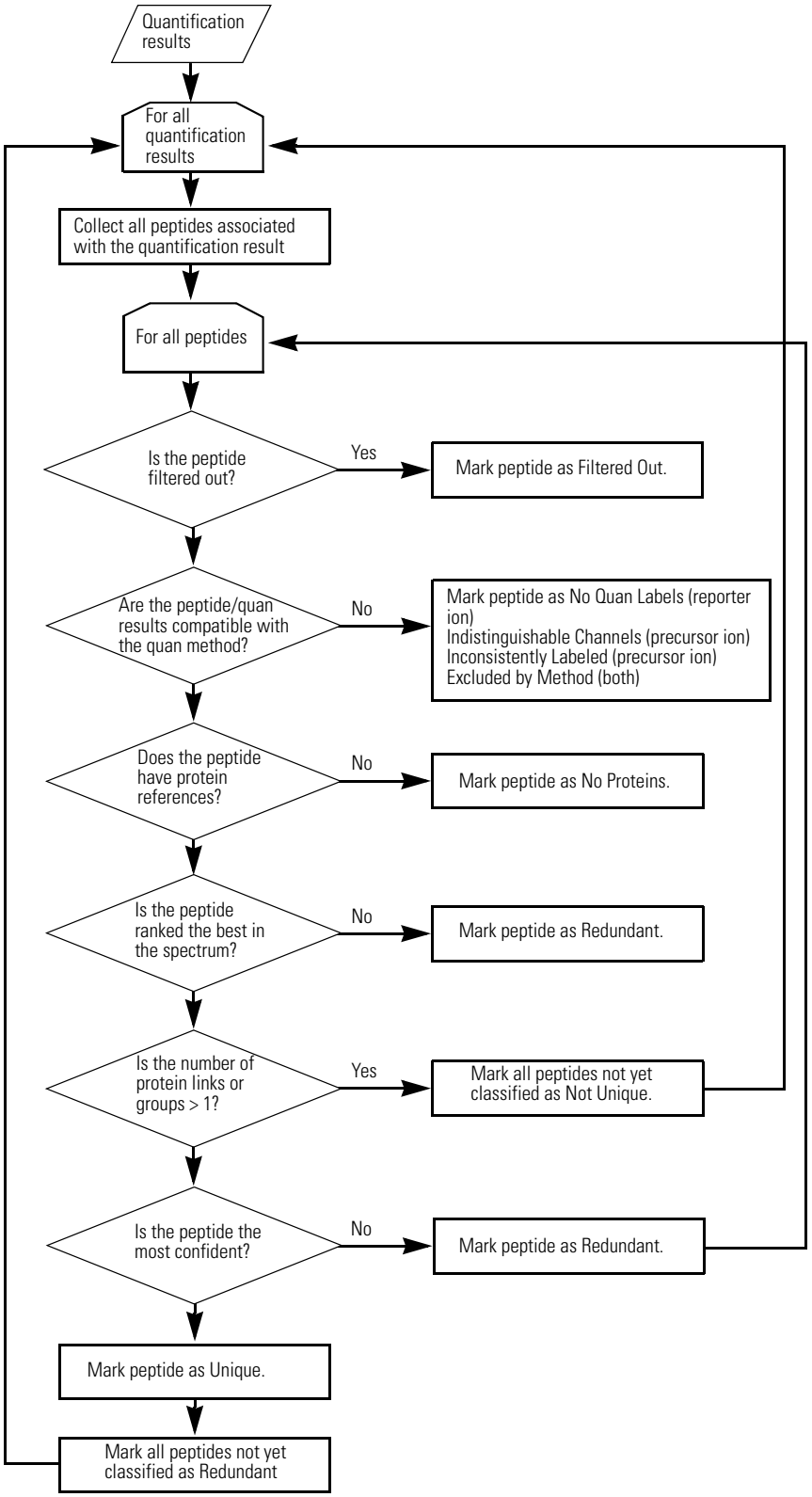
Note The Proteome Discoverer application only calculates the % Isolation Interference value if the precursor scans are high-resolution, high-mass-accuracy scans.

You can use the Percentage Co-Isolation Excluding Peptides from Quantification parameter on the Ratio Calculation page of the Quantification Method dialog box (shown in Figure 278 on page 536) to specify a threshold of between 0 and 100 percent for the allowed co-isolation interference. The default value is 100 percent, which means that no PSM is excluded. This parameter is only available for reporter ion quantification.

Classifying Peptides

The chart in Figure 285 shows how the Proteome Discoverer application classifies peptides for protein quantification. It displays this classification in the Quan Info column of the results report. These classifications are described in "Peptides Page Parameters" on page 291.

Figure 285. Classifying peptides for protein quantification



Depending on the settings in the Quantification Method Editor, the Proteome Discover application excludes the following peptides:
- No Quan Labels: No reporter label
- Indistinguishable Channels: Not all defined channels can be distinguished
- Inconsistently Labeled: Labels are from different channels
- Excluded by Method: Quan channels are missing, single-peak channels are missing, ratios exceed limits, and so forth

The Proteome Discoverer application does not consider high-scoring peptides if they have no protein links.

The Proteome Discoverer application marks peptides that are not unique as Redundant.

This check depends on the setting of the Consider Protein Groups for Peptide Uniqueness check box on the Protein Quantification page of the Quantification Method Editor dialog box. It classifies the peptides as Unique if they match the proteins within the same protein group.

Calculating Peptide Ratios

For both precursor ion and reporter ion quantification, the Proteome Discoverer application calculates protein ratios as the median, not the mean, of all peptide hits belonging to a protein that is marked “Used” in the Quan Usage column of the report. It chooses the median to calculate the protein ratios because it is relatively robust in the presence of outliers. In principle, the Proteome Discoverer application uses only the peptides in the filtered results for protein ratio calculation when the result filters are applied to the search result. These result filters are what you want to apply to quantification. For example, protein ratios that change because you filter peptides having a specific sequence tag will skew the results.

Protein ratios are the median of the peptides of the protein. If you want to recalculate the peptide ratio, you must ensure that all peptides are displayed. By default, the application considers only unique peptides in the calculation so that only peptides that have no other protein references are considered.

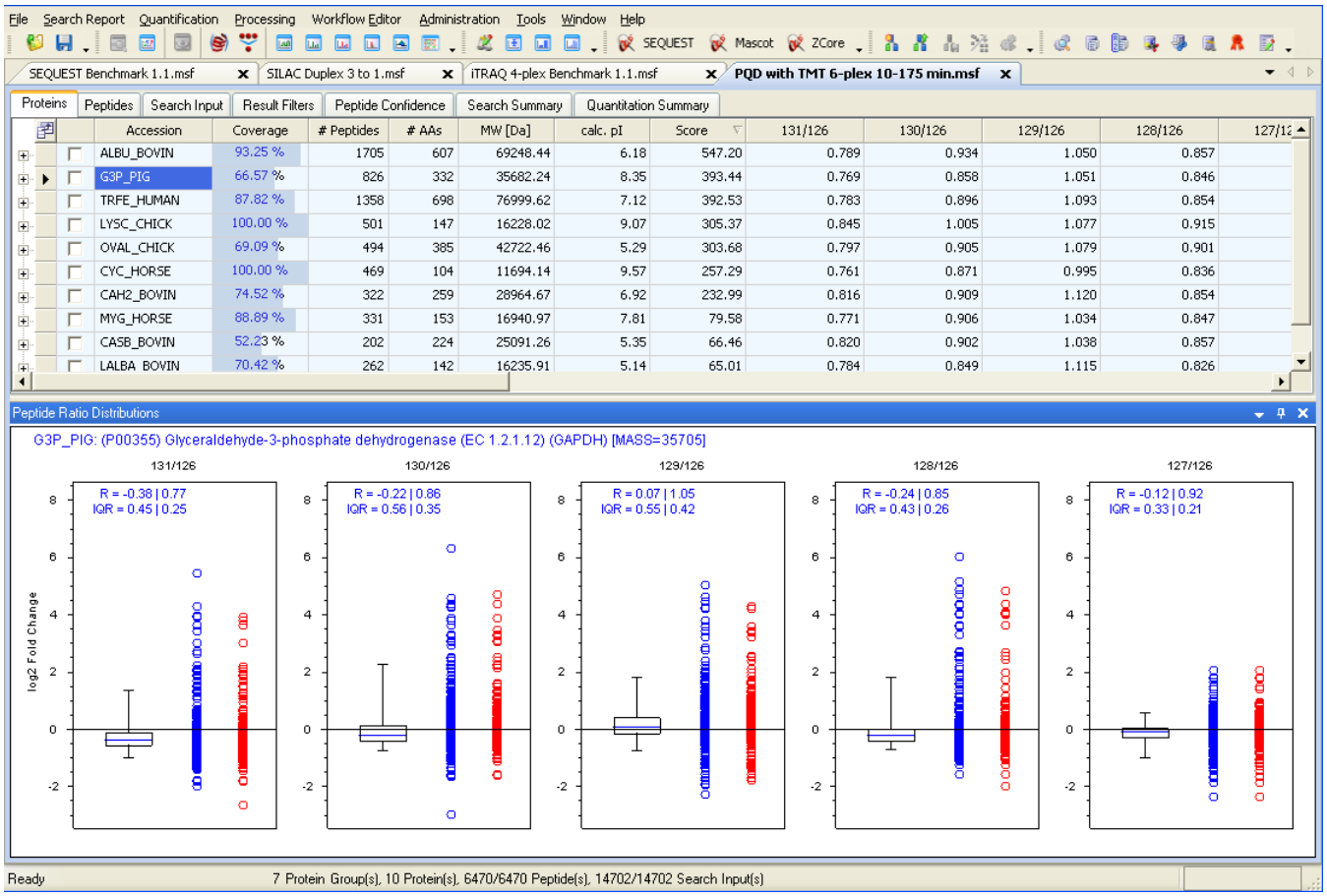
You can activate a chart of the peptide ratios. This graph shows the distribution of peptide ratios for the selected protein, displaying the ratios of the peptides associated with the selected protein as a log2-fold change.

❖ To calculate peptide ratios

1. Click the row of the peptide or protein that you are interested in.
2. Choose **Quantification > Show Peptide Ratios**, or click the **Show Peptide Ratios** icon,  .

The Peptide Ratio Distributions chart shown in Figure 286 appears. The following sections describe the pages available in this view.

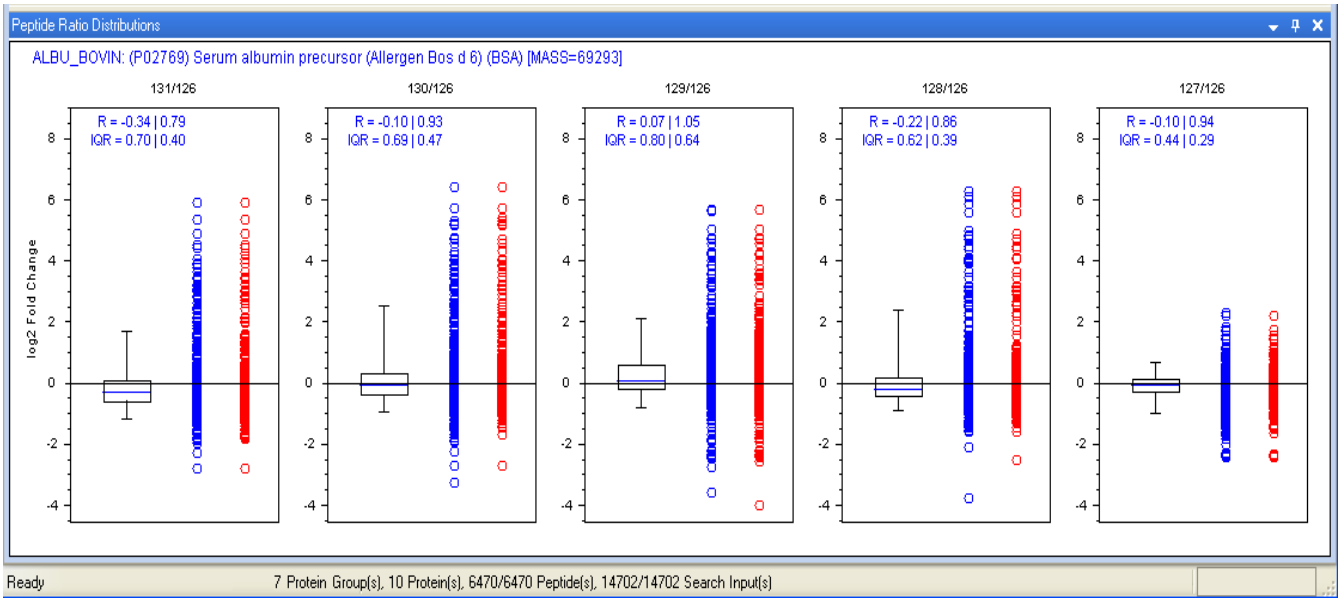
Figure 286. Peptide Ratio Distributions chart



Understanding the Peptide Ratio Distributions Chart

The Peptide Ratio Distributions chart shows the distribution and spread of the ratios of all peptides belonging to a particular protein. Figure 287 shows an example for the albumin protein.

Figure 287. Peptide Ratio Distributions chart



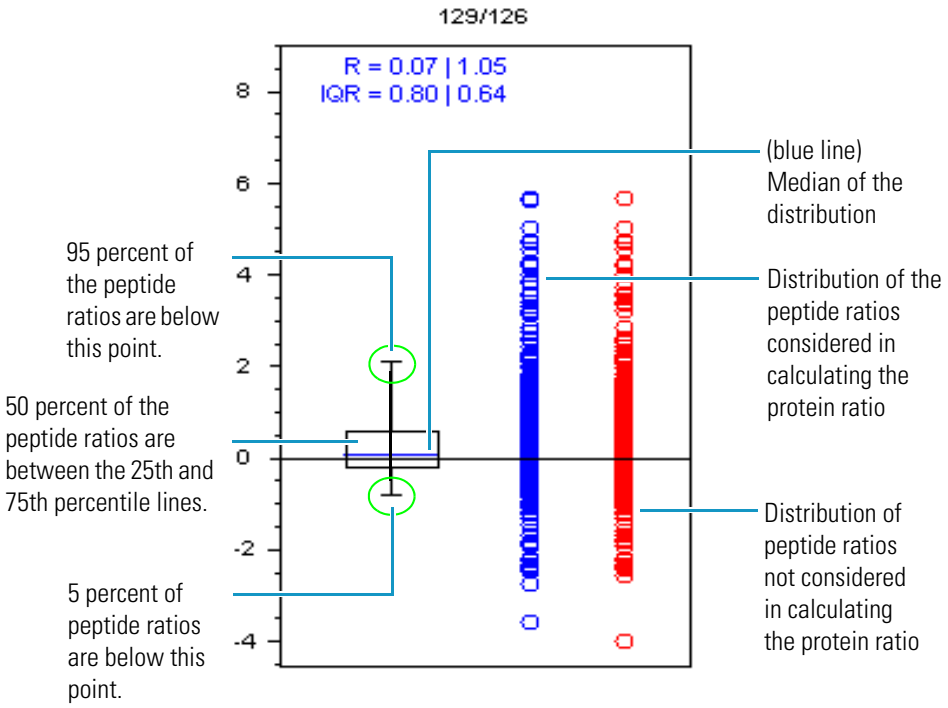
The chart shows the distribution of peptide ratios for each of the ratios reported, as defined in the quantification method for this search. Each of the ratio distribution charts displays the peptide ratios as the binary logarithm. The logarithmic form is common for such displays, because it provides a reasonable display, even when there is a large spread of the displayed values. In binary logarithmic form, a value of 1 means a two-fold increase, a value of 2 means a four-fold increase, a value of 3 means an eight-fold increase, and so forth. Each of the separate distribution charts displays the peptide ratios in three sections. The chart legend explains the meaning of these sections. You can access the chart legend by right-clicking the chart and choosing Show Legend.

The Peptide Ratio Distribution charts contain the three sections illustrated in Figure 288:

- The first section displays the distribution of the ratios of all peptides considered for calculating the ratio of this protein as a box-and-whisker plot. A box-and-whisker plot is a convenient way of graphically depicting groups of numerical data through a five-number summary: 5 percent lower bound, lower quartile, median, upper quartile, 95 percent upper bound. The range between the lower and upper quartile (this is the range of the box) is also known as the inter-quartile range (IQR) and, like the standard deviation for normally distributed data, is a measure of the spread of the data.
  - The box represents the peptide ratios between the 25th and the 75th percentiles.
  - The error bars represent the peptide ratios below the 5th and the 95th percentiles.
  - The blue lines inside the horizontal bar represent the median of the distribution.
- The second section (blue circles) displays the distribution of the ratios of all peptides considered in calculating the protein ratio.

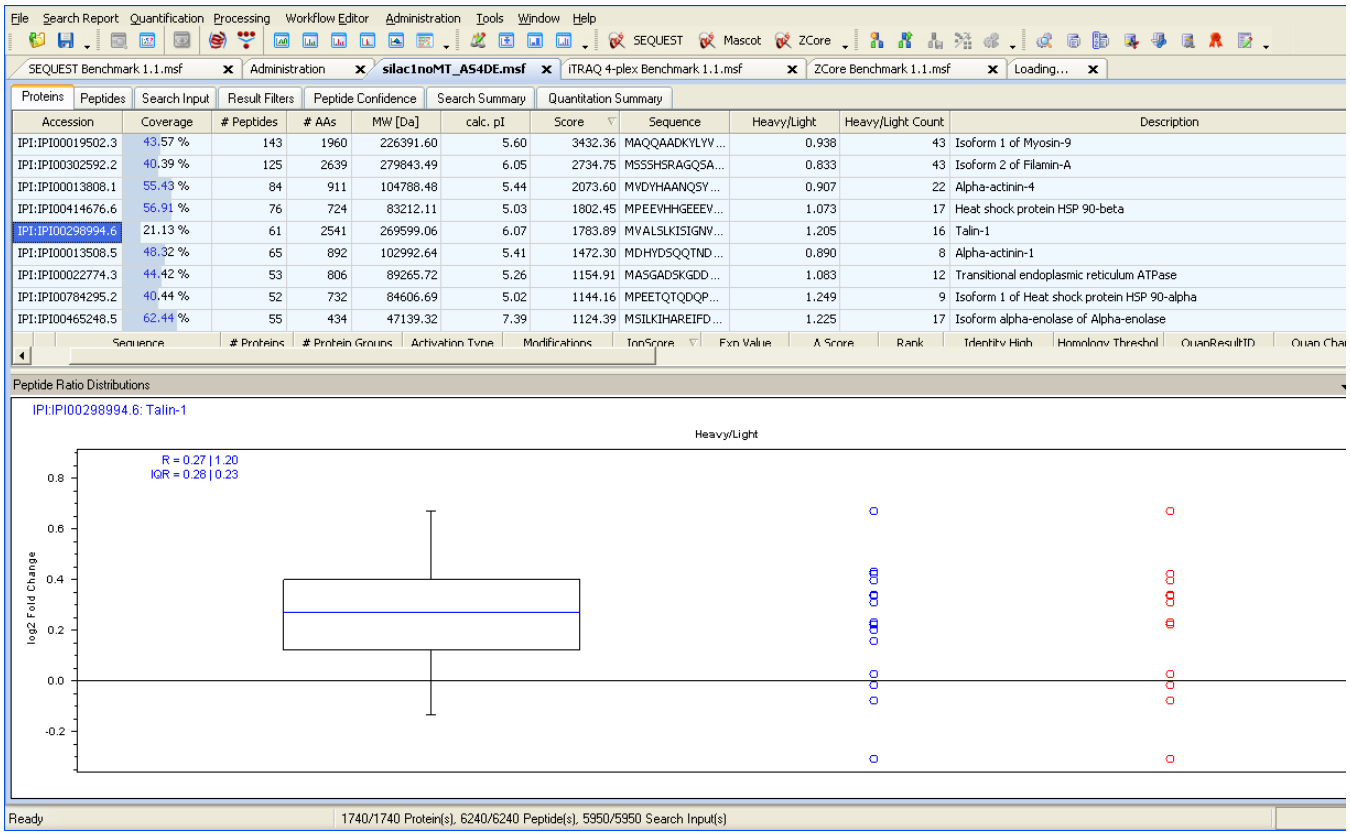
- The third section (red circles) displays the distribution of the ratios of all peptides that were not considered in calculating the protein ratio (for example, the peptide ratio was considered too extreme, or this peptide is not unique to this protein or this protein group) according to the rules defined in the quantification method.

Figure 288. Peptide Ratio Distributions chart for reporter ion quantification



In addition, each chart displays the median ratio (R) and the inter-quartile range (IQR) in linear and logarithmic format.

Figure 289. Peptide Ratio Distributions Chart for precursor ion quantification



The header of the chart identifies the protein that the peptide belongs to. Right-click the chart and choose **Show Legend** for the identity of other notations on the chart.

Handling Missing and Extreme Values in Calculating Peptide Ratios

Table 93 and Table 94 list some of the different circumstances that can arise in calculating quantification ratios for peptides from the selected quantification values. A quantification value is the intensity or area detected for a given quantification channel. For reporter ion quantification, a quantification channel is one of the mass or reporter tags, and for precursor ion quantification, it is one of the different possible labeling states of a peptide corresponding to the different heavy amino acids used in the cell cultures. “Intensity” refers to both the intensity of the reporter peaks in reporter ion quantification and to the areas detected in precursor ion quantification.

When the Proteome Discoverer application detects the quantification values for the different quantification channels, some of the quantification values might be missing, probably because they fell below the detection limit. In addition, some channels might show very low or very high intensities, leading to the calculation of very high or very low ratios. Major changes might indicate exceptional cases, which you can exclude from the calculation of the protein ratios by using the settings on the Ratio Calculation page of the Quantification Method Editor dialog box.

Table 93 and Table 94 do not include cases that arise as a result of peptide uniqueness and protein grouping. They focus on cases that arise where one or both of the quantification channels that are used for calculating peptide ratios are zero. In these cases, the application detects nothing on a channel because the spectrum does not contain one of the reporter peaks, the heavy or light isotope pattern is missing, a quantification value falls below a specified minimum threshold, or the calculated ratios are very high or very low.

Table 93 and Table 94 list the different possible cases exemplified by arbitrary values. The values in the tables have [counts] as units if the cases are presented for reporter ion quantification. For precursor ion quantification, 114 and 115 are replaced by Light and Heavy, and the quantification values have [counts × min] as units.

In addition to the options listed in the tables, the handling of quantification values is also affected by the Apply Quan Value Corrections option on the Ratio Calculation page of the Quantification Method Editor dialog box and by the options on the Experimental Bias page of the same dialog box. For reporter ion quantification, the Apply Quan Value Corrections option determines whether to apply the purity correction for the detected quantification values. The Proteome Discoverer application applies the purity correction after it applies the other settings that potentially change the quantification values. It applies the experimental bias correction after the first time that it calculates all peptide and protein ratios. The application then determines the bias correction factor and applies it to every peptide and protein ratio.

Table 93. Calculating peptide ratios when quantification values are missing (Sheet 1 of 2)

Case	Minimum detected quan. value	Minimum Quan Value Threshold setting	Replace Missing Quan Values with Minimum Intensity setting	Reject All Quan Values If Not All Quan Channels Are Present setting	Detected quantification values			Displayed/used quantification values		
					114	115	116	114	115	116
All quan. values detected	33	0	No	Irrelevant	100	50	300	100	50	300
	33	0	Yes	Irrelevant	100	50	300	100	50	300
	33	75	No	Irrelevant	100	50	300	100	0	300
	33	75	Yes	Irrelevant	100	50	300	100	75	300

Table 93. Calculating peptide ratios when quantification values are missing (Sheet 2 of 2)

Case	Minimum detected quan. value	Minimum Quan Value Threshold setting	Replace Missing Quan Values with Minimum Intensity setting	Reject All Quan Values If Not All Quan Channels Are Present setting	Detected quantification values			Displayed/used quantification values		
					114	115	116	114	115	116
Quan. value missing for a quan. channel	33	0	No	No	100	0	300	100	0	300
	33	0	Yes	No	100	0	300	100	33	300
	33	75	No	No	100	0	300	100	0	300
	33	75	Yes	No	100	0	300	100	75	300
	33	0	No	Yes	100	0	300	0	0	0
	33	0	Yes	Yes	100	0	300	100	33	300
	33	75	No	Yes	100	0	300	0	0	0
	33	75	Yes	Yes	100	0	300	100	75	300
Quan. value missing for all quan. channels	33	0	No	Irrelevant	0	0	0	0	0	0
	33	0	Yes	Irrelevant	0	0	0	0	0	0
	33	75	No	Irrelevant	0	0	0	0	0	0
	33	75	Yes	Irrelevant	0	0	0	0	0	0

Table 94. Calculating peptide ratios when values are very high or low (Sheet 1 of 2)

Case	Maximum Allowed Fold Change setting	Use Ratios Above Maximum Allowed Fold Change for Quantification setting	Calculated ratios		Displayed ratios	
			115/114	116/114	115/114	116/114
Ratio is within the limits	100	Irrelevant	2.000	3.000	2.000	3.000
	100	Irrelevant	0.500	0.250	0.500	0.250
Ratio is 0 or ∞ because one quan. channel value is missing	100	No	0	3.000	0.000	3.000
	100	Yes	∞	0.250	100.000	0.250

Table 94. Calculating peptide ratios when values are very high or low (Sheet 2 of 2)

Case	Maximum Allowed Fold Change setting	Use Ratios Above Maximum Allowed Fold Change for Quantification setting	Calculated ratios		Displayed ratios	
			115/114	116/114	115/114	116/114
A ratio exceeds the limits	100	No	2000.000	3.000	100.000	3.000
	100	No	0.300	0.002	0.300	0.010
	100	Yes	2000.000	3.000	100.000	3.000
	100	Yes	0.300	0.002	0.300	0.010
All ratios exceed the limits	100	No	2000.000	0.002	100.000	0.010
	100	Yes	2000.00	0.002	100.000	0.010

Calculating Protein Ratios from Peptide Ratios

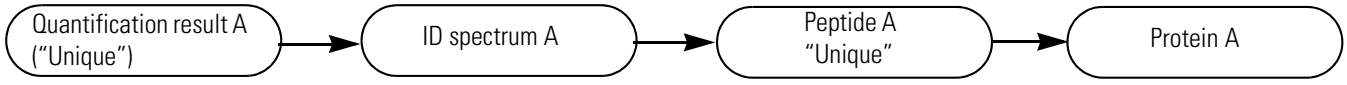
This section describes seven different scenarios that can occur when you derive protein quantification ratios from peptide quantification ratios. These cases show how the validity of using a given quantification result for the quantification of a certain protein depends on whether this particular quantification result is unique or shared among other peptides.

The peptide quantification ratios are taken from the associated quantification results. The term *quantification result* in this section refers to MS/MS reporter intensities taken from the same scan as the identification (for example, ID-CID) or from a separate quantification scan (for example, Quan-HCD). The term also refers to intensities derived from the precursor scans in precursor ion quantification. A quantification result here is a general quantity associated with one or more peptides that are, in turn, associated with one or more proteins.

Case 1: Quantification Result Associated with One Spectrum, One Peptide, and One Protein

Case 1, shown in Figure 290, is the simplest case. The quantification result is associated with one identification spectrum—whether the quantification results come from the same identification spectrum, from a different quantification spectrum, or from the precursor ion—and one peptide that is contained in one protein. The quantification result is unique for this protein. The Proteome Discoverer application can mark peptide A "Unique" in the Quan Info column of the Peptides page if the quantification result meets other criteria.

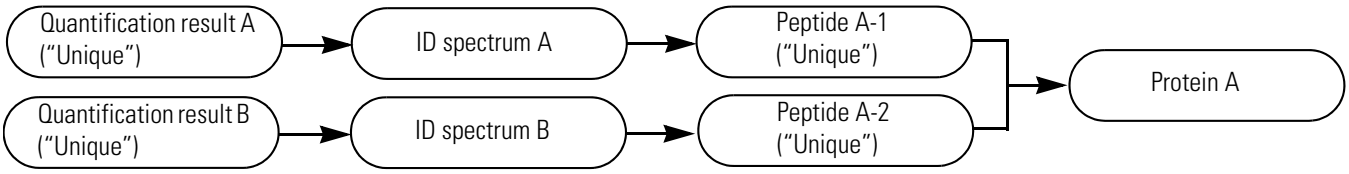
Figure 290. Case 1: Quantification result associated with one identification spectrum, one peptide, and one protein



Case 2: Two Quantification Results Associated with Two Spectra, One Peptide, and One Protein

Case 2, shown in Figure 291, is a variant of case 1. Each of two different quantification results is associated with a different identification spectrum. Both identification spectra identify peptide A, which is a peptide with the same sequence. Peptide A is only contained in one protein. Each of the two different quantification results is unique for just one protein. The peptides are redundantly identified and quantified, and you could use both for the quantification of protein A.

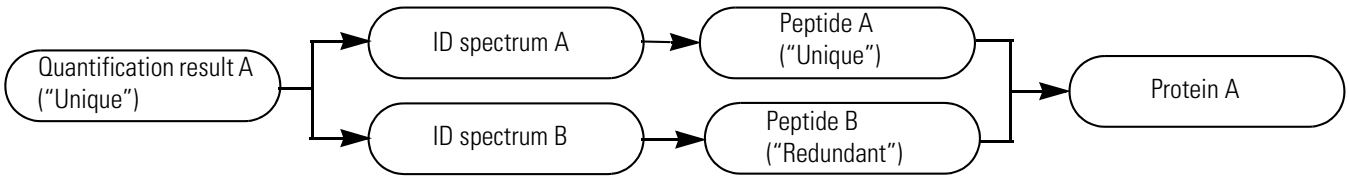
Figure 291. Case 2: Two different quantification results associated with two identification spectra, one peptide, and one protein



Case 3: Quantification Result Associated with Two Spectra, Two Peptides, and One Protein

Case 3, shown in Figure 292, is similar to case 2 but varies from it in a slight but important way. In case 3, the two identification spectra are associated with the same quantification result rather than with two different quantification results. For example, you might obtain these results if you trigger the same precursor two times for MS/MS. It does not matter whether peptide A and peptide B are the same peptides (redundantly identified) or different peptides that are accidentally contained in the same protein. It also does not matter whether they are identified by the same search engine or by two different search engines, for example, a CID spectrum and an ETD spectrum. The quantification result is still unique for just one protein. However, you cannot use the quantification ratio of both peptides A and B to calculate the quantification ratio of protein A, because it is the same quantification result, and you do not want to use the same quantification result multiple times for the same protein. In this case, the Proteome Discoverer application marks peptide A, the peptide with the better identification, as "Unique" and the other peptide as "Redundant" for quantification (rather than redundant for identification).

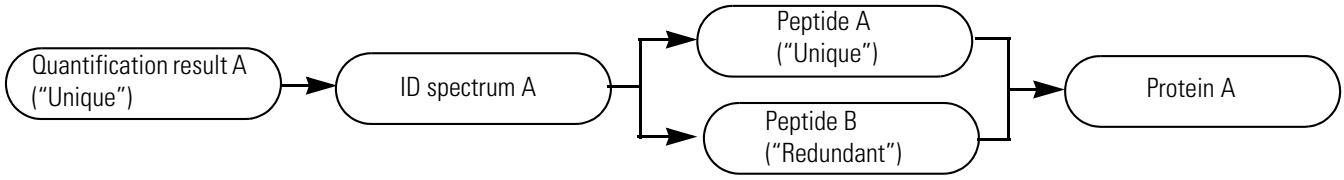
Figure 292. Case 3: Quantification result associated with two identification spectra, two peptides, and one protein



Case 4: Quantification Result Associated with One Spectrum, Two Peptides, and One Protein

In case 4, shown in Figure 293, the two peptides could be identified by the same search engine and have different ranks, or they could be identified by different search engines and both have rank 1. It does not matter whether peptide A and B have the same sequence with different PTM states or different sequences. The quantification result is unique for protein A. You can use it to calculate the protein ratio, but you must only use it once. The Proteome Discoverer application marks the "better" peptide as "Unique" and the other as "Redundant" for quantification.

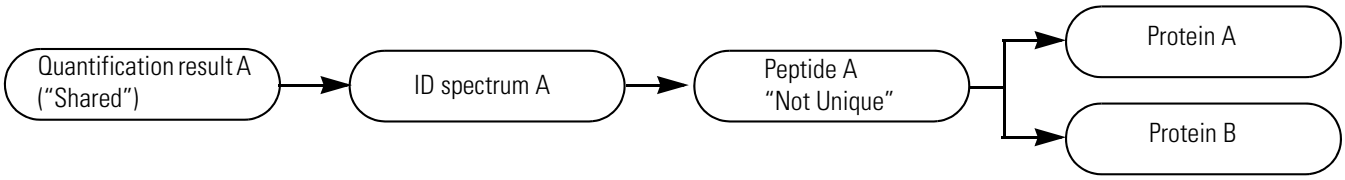
Figure 293. Case 4: Quantification result associated with one identification spectrum, two peptides, and one protein



Case 5: Quantification Result Associated with One Spectrum, One Peptide, Two Proteins

In case 5, shown in Figure 294, the quantification result is associated with one identification spectrum and one peptide, but this peptide is contained in more than one protein. The quantification result is potentially shared between these proteins, and you do not know how to share it. If the quantification method specifies using only unique peptides for protein quantification, you would not use peptide A in this case. If the quantification method specifies using all peptides for protein quantification, the quantification result of peptide A would be divided equally between both proteins.

Figure 294. Case 5: Quantification result associated with one identification spectrum, one peptide, and two proteins

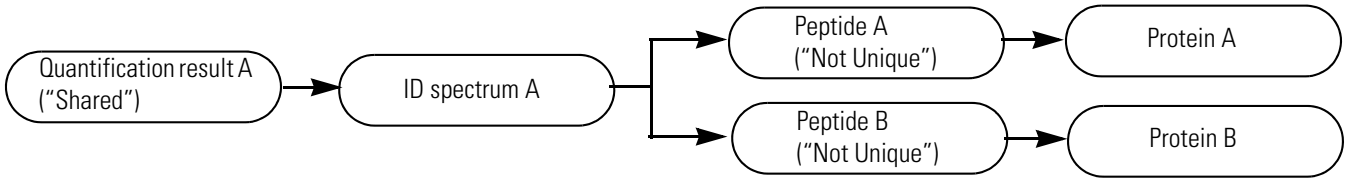


Case 6: Quantification Result Associated with One Spectrum, Two Peptides, and Two Proteins

In case 6, shown in Figure 295, the quantification result is associated with one identification spectrum from which two different peptides are identified either by the same search engine as different ranks or by different search engines. The two different peptides are contained in two different proteins. The two different peptides are both unique to just one protein. Nevertheless, the associated quantification result is the same, and you do not want to use it for

the calculation of the protein ratios if you specified in the quantification method to use only unique peptides. Only if you specify using all peptides can you use them for protein quantification. This case illustrates the discrepancy between the uniqueness of peptides and the uniqueness of the quantification results.

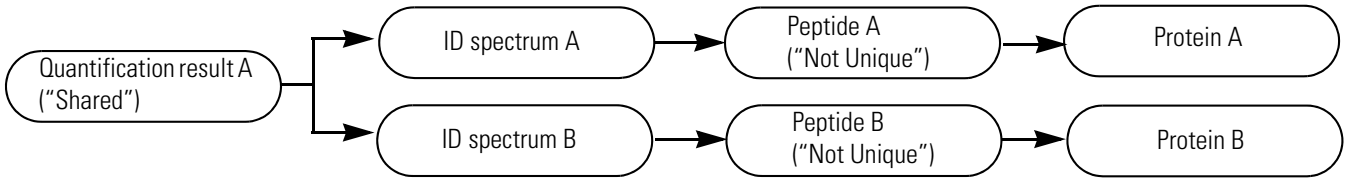
Figure 295. Case 6: Quantification result associated with one identification spectrum and two peptides unique to one protein



Case 7: Quantification Result Associated with Two Spectra, Two Peptides, and Two Proteins

Case 7, shown in Figure 296, is a variant of case 6. Either the same search engine or different search engines identify different identification spectra, for example, CID and ETD spectra. As in case 6, the peptides are unique, but the quantification result is not. The result depends on whether you specified in the quantification method to use only unique peptides or all peptides.

Figure 296. Case 7: Quantification result associated with two identification spectra and two different peptides unique to one protein



Calculating Ratio Count and Variability

The Proteins page of search reports with precursor ion quantification results displays columns called Heavy/Light Variability and Heavy/Light Count. Similarly, the Proteins page of search reports containing reporter ion quantification results displays columns called *Ratio* Variability [%] (for example, 114/113 Variability [%]) and *Ratio* Count (for example, 114/113 Count). The way the Proteome Discoverer application calculates and displays the values in these columns depends on whether you want the results treated as replicates or as treatments.

Replicates

Replicates are repeated measurements of the same sample. You repeat measurements to obtain better statistics. Without replicates, you cannot be sure that something that you observed is real—that is, statistically significant—and not a result of an error in the sample preparation, the liquid chromatography, the acquisition, and so forth. To generate replicates, you can

repeat the sample preparation or use the same sample and measure it multiple times. This data highlights the variance within the different steps. For example, if you measure a difference of 17 percent between two samples representing different treatments with a new drug, but the variance between the different replicates of the same treatment is already 28 percent, the observed difference might not be significant.

Treatments

Treatments are samples that are brought to different states. For example, they might be different samples representing different exposure levels to a certain drug or cultures of the same cells exposed to different levels of stress, such as radiation, salts, or heat.

Ratio Count

The *Ratio* Count or the Heavy/Light Count column displays the number of peptide ratios that were used to calculate a particular protein ratio. If only one ratio was reported (for example, the Heavy/Light ratio for SILAC data), the displayed count is the number of peptides marked “Used” for this protein. If more than one ratio was used (for example, the ratios in iTRAQ or TMT data), the count for a particular protein ratio might be smaller than the number of peptides marked “Used.” The count could be lower than the number of peptides marked “Used” in cases where some of the ratios are excluded by the different settings or thresholds defined by the quantification method.

For replicates, the *Ratio* Count columns display a list of the separate counts for each replicate. If a protein was not identified in one of the replicates or no peptide usable for calculating the protein was identified for this replicate, a “-” appears in the *Ratio* Count cell. If none of the replicates provide a usable peptide, the *Ratio* Count cell is empty.

Ratio Variability

The *Ratio* Variability [%] columns show the variability of the peptide ratios that are used to calculate a particular protein ratio. They are similar to a coefficient of variation for the calculated protein ratios as a normalized measure of the peptide ratio spread used for calculating the protein ratio. The Proteome Discoverer application calculates the displayed variability differently for single search reports, multiconsensus reports that are treated as treatments, and multiconsensus reports that are treated as replicates.

Single Search Reports

For single search reports, the protein ratio variability is calculated as a coefficient-of-variation for log-normal distributed data ($CV_{log\text{-}normal}$). In this case, the protein ratio variability is calculated from the used peptide ratios $r_1 \ldots r_n$ as follows:

$$CV_{log_normal} = \sqrt{\exp([StdDev(\log(r_1)\ldots\log(r_n))]^2) - 1}$$

where $StdDev\ (\log(r_1) \ldots \log(r_n) = 1.483 \times MAD(\log(r_1) \ldots \log(r_n))$.

$$CV_{log_normal} = \sqrt{\exp([1.483 \times MAD(\log(r_1)\ldots\log(r_n))]^2) - 1}$$

$$variability(r_1\ldots r_n) = 100 \times CV_{log\text{-}normal}$$

where $MAD(r)$ is the median absolute deviation (MAD) of the peptide ratios $r_1 \ldots r_n$. In statistics, the median absolute deviation is a robust measure of the variability of a univariate sample of quantitative data.

$$MAD(\log(r_1 \ldots r_n)) = median[|\log(r_1) - median(\log(r_1 \ldots r_n))|]$$

Starting with the residuals (deviations) from the data's median, the median absolute deviation is the median of their absolute values. The 1.483 constant ensures consistency for the r_i distributed normally as $N(\mu, \sigma 2)$ and large N:

$$E(1.483 \times MAD(r_1\ldots r_n)) = \sigma$$

The Proteome Discoverer application uses these statistics because they are more robust in the presence of outliers as a classical coefficient of variation (CV). It also uses them to calculate the protein ratio as the median of the used peptide ratios.

Calculating Variability in Multiconsensus Reports Treated as Treatments

For multiconsensus reports that treat quantification data as different treatments, the results of the single searches are simply displayed side by side, and the variabilities are the same as those of the single reports. For more information on how the Proteome Discoverer application calculates protein ratios when treating quantification results in multiconsensus reports as treatments, see "Calculating Protein Ratios in Multiconsensus Reports Treated as Treatments" on page 565.

Calculating Variability in Multiconsensus Reports Treated as Replicates

For multiconsensus reports that treat quantification data as replicates, the Proteome Discoverer application calculates the protein ratios for single searches and then calculates a classical coefficient variation for these ratios. It calculates the variability of the protein ratio calculated from N replicates from the protein ratio $r_1 \ldots r_n$ of the single searches:

$$variability(r_1 \ldots r_n) = CV = 100 \times \frac{\text{std. dev.}(r_1 \ldots r_n)}{\text{arith. mean}(r_1 \ldots r_n)}$$

Using the protein ratios rather than their logarithms is reasonable because in contrast to the peptide ratios, which are (at least approximately) log-normally distributed, the protein ratios of the single searches should be normally distributed—at least for larger values of n. For more information on how the Proteome Discoverer application calculates protein ratios when treating quantification results in multiconsensus reports as replicates, see "Calculating Protein Ratios in Multiconsensus Reports Treated as Replicates" on page 566.

Calculating and Displaying Protein Ratios for Multiconsensus Reports

The Proteome Discoverer application can treat the different single quantification results in multiconsensus reports as replicates of the same sample or as different treatments of a sample.

In real-world studies of quantitative responses of a sample to certain treatments, such as a particular change in environmental condition or an administration of a drug, you might be interested in the quantitative difference of the sample before and after the treatment, or between different treatment states of the sample (for example, different points in time after application of a certain drug or application of different amounts of a certain drug). Quantitative studies could also investigate the quantitative difference between samples in different states, for example, between similar samples from healthy and different disease states.

Such experimental investigations must assess the variability inherent in the different stages of the experiment. For example, samples from different animals or patients can vary significantly in their expression level for certain proteins or in the amount of proteins and peptides with PTMs. Other sources of variability are differences in sample preparation, differences in chromatographic separation, or differences in measurement in the mass spectrometer. When you examine the quantitative differences between two measurements, all these single factors combine to create an overall variability of the quantitative values under investigation, for example, the expression levels of certain proteins.

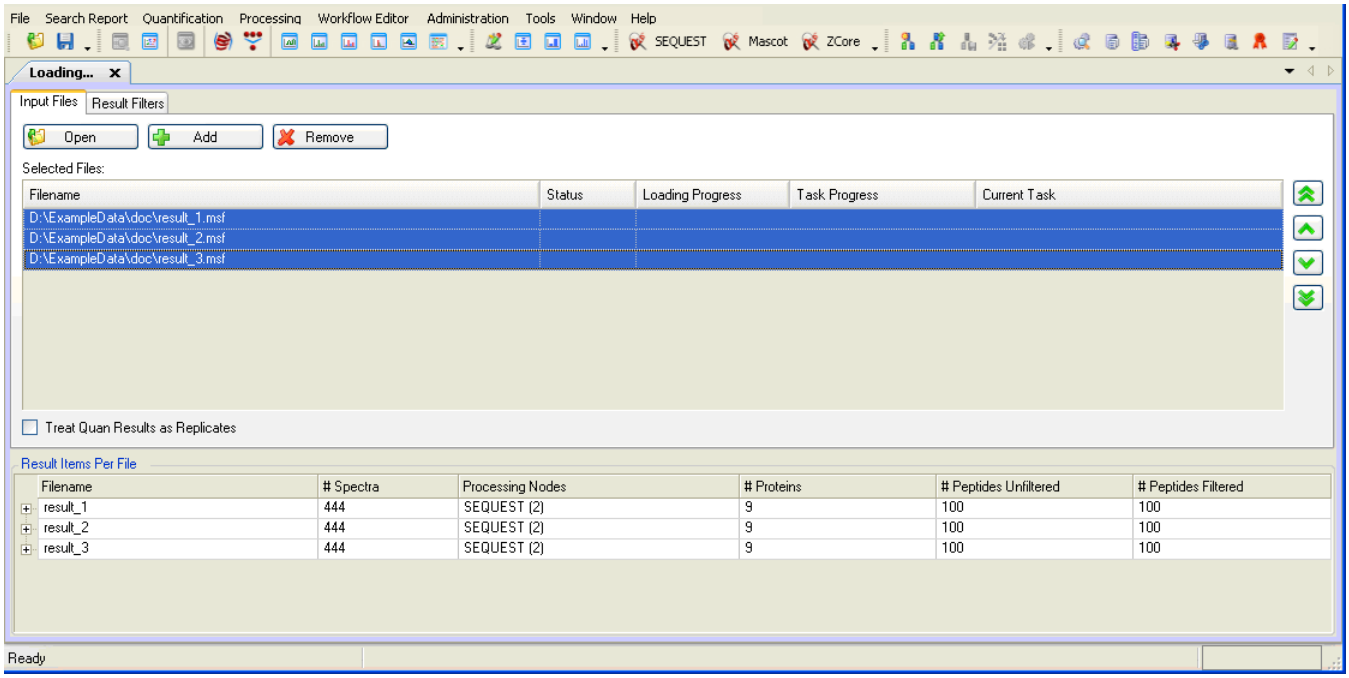
This overall variability can be quite significant. To minimize the variability when comparing two samples, such as different treatments or disease states, and to calculate a statistical measure of the inherent variability, you must measure replicates. In this process, you repeatedly measure a sample multiple times and calculate the average values for the

quantitative values under investigation. You perform these measurements for all states of the sample and then compare the calculated average values. You can then calculate whether a detected difference between two states of a sample is statistically significant or is only due to the inherent variability of the sample.

In the Proteome Discoverer application, you can load multiple result files containing quantification results and treat the single results as replicates of the same sample or as different treatments of a sample. You determine whether the single results of an open multiconsensus report should be treated as replicates or treatments, and you can change them from replicates to treatments and vice-versa.

As an example, assume that you have three result files from measuring and processing a yeast sample: result_1.msf, result_2.msf, and result_3.msf. Assume that the samples were prepared with the iTRAQ 4plex quantification method, giving quantifiable reporter peaks at 114, 115, 116, and 117 *m/z*. When you open these three MSF files, the Proteome Discoverer application adds the files to the Input Files page, as shown in Figure 297.

Figure 297. Loading three single result files containing quantification data



By default, the Proteome Discoverer application initially treats the quantification results in the single MSF files as if they were from different treatments of a sample. You can change this treatment by selecting the Treat Quan Results as Replicates option on the Input Files page. The application then initially treats the quantification data as if it were replicates of the same sample when it creates and opens the multiconsensus report.

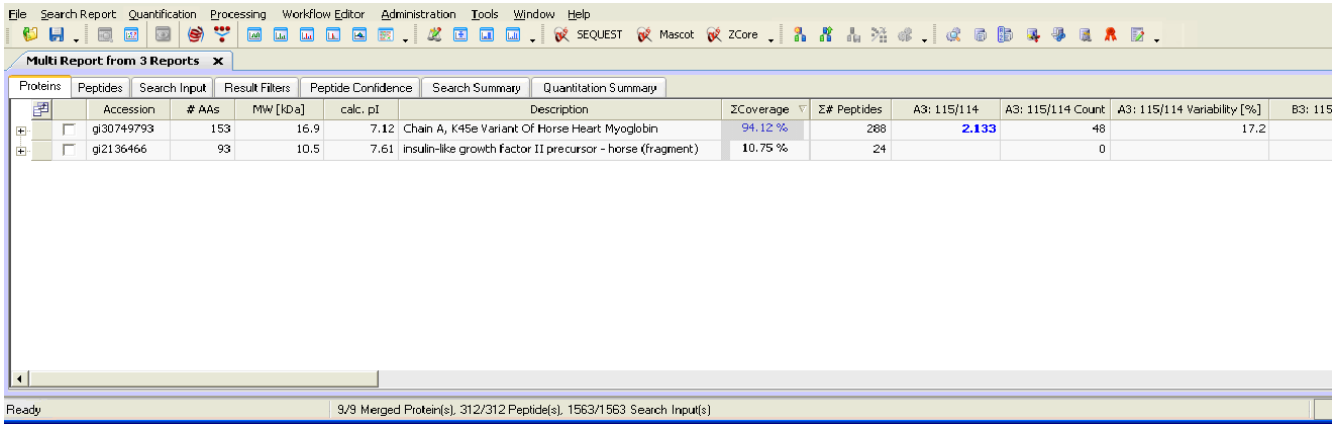
If you do not select the Treat Quan Results as Replicates option and click Open on the Input Files page, the application creates the multiconsensus report and calculates the quantification results (the quantification ratios as specified in the quantification method) for each of the three single results files separately. In this example, all three result files are iTRAQ 4plex files, so the application usually calculates ratios such as 115/114, 116/114, and 117/114 for each of the files. Because you did not select Treat Quan Results as Replicates, the application reports them as if the three files represented different treatment states of a sample.

Calculating Protein Ratios in Multiconsensus Reports Treated as Treatments

In the case of treatments, the Proteome Discoverer application reports all calculated protein ratios of the single result files side by side for the multiconsensus report. It prefixes each file by a single letter identifier of the particular report and the number of the quantification node in the processing workflow. In the example given, the protein and corresponding variability and ratio count columns are A3: 115/114, B3: 115/114, C3: 115/114, A3: 116/114, B3: 116/114, C3: 116/114, A3: 117/114, B3: 117/114, and C3: 117/114, as shown in Figure 298. At the peptide level, there is no difference between treatments or replicates in multiconsensus reports and single reports.

Figure 298 shows the protein ratios in a multiconsensus report when the quantification results of the single result files are treated as different treatments of the sample.

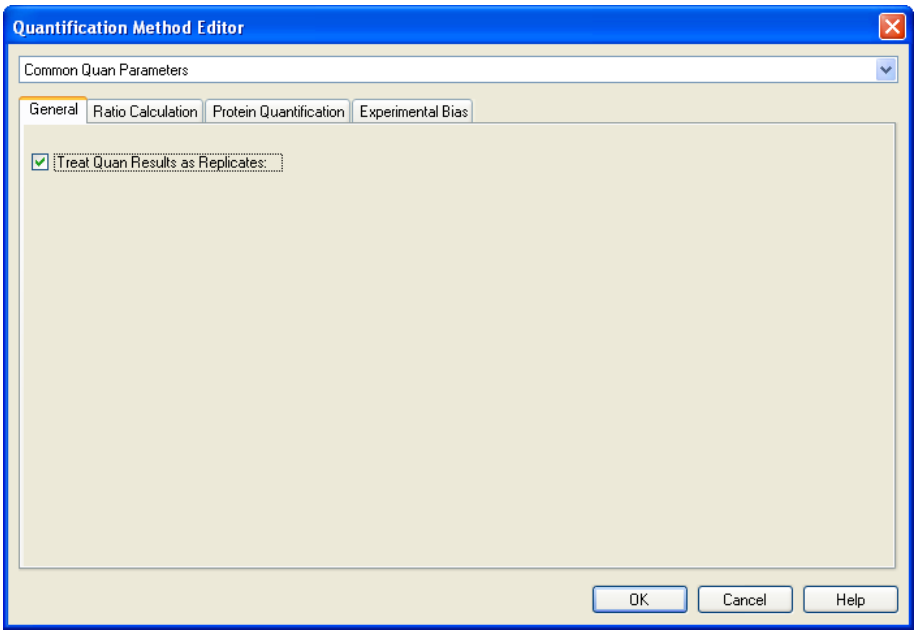
Figure 298. Protein ratios when single quantification result files are treated as different treatments of the sample



Calculating Protein Ratios in Multiconsensus Reports Treated as Replicates

With replicates, the application treats the quantification results like replicates of the same sample. You can specify that quantification results be treated as replicates by selecting the Treat Quan Results as Replicates option on the Input Files page or by using the Quantification Method Editor dialog box (opened by choosing Quantification > Edit Quantification Method) when the multiconsensus report is open. For multiconsensus reports, the Quantification Method Editor dialog box features a Common Quan Parameters box (shown in Figure 263 on page 503) so that you can set common quantification parameters for all contained result files at once. On the General page of the dialog box, shown in Figure 299, you can switch between treatment and replicate mode.

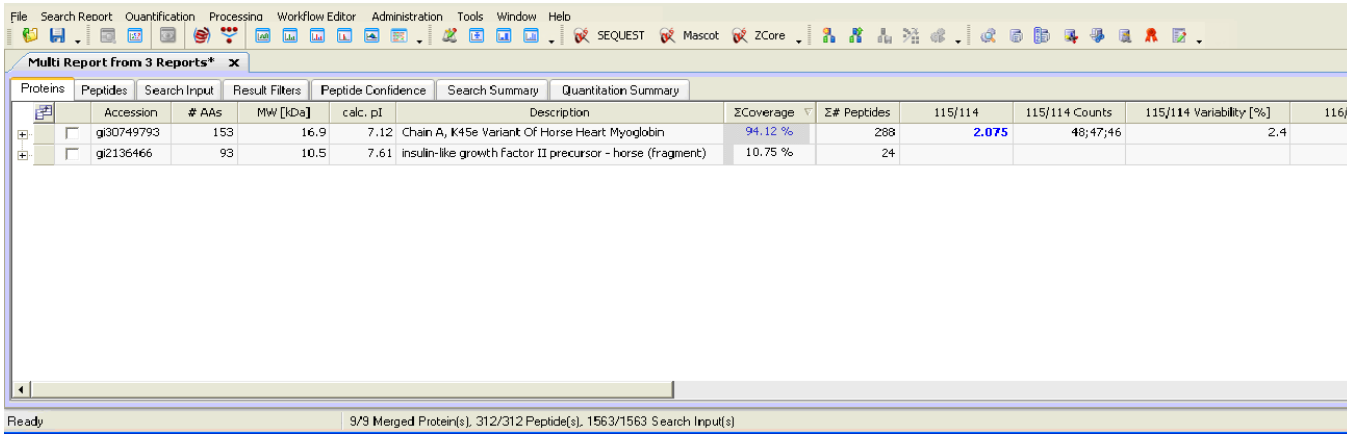
Figure 299. Switching between treatment and replicate mode by editing the common quantification parameters



As in treatment mode, multiconsensus reports are no different from single reports for replicates at the peptide level. At the protein level, the Proteome Discoverer application combines the protein ratios of the single result files into averaged protein ratios, as shown in Figure 300. It calculates the combined protein ratio as the arithmetic mean of the protein ratios of the single reports (and calculates the protein ratios as the median of the "used" peptide ratios of the particular result file). See "Calculating Ratio Count and Variability" on page 560 for information on how the application calculates and displays the values in the Ratio Count and Variability columns for multiconsensus reports.

Figure 300 shows protein ratios in a multiconsensus report when the quantification results of the single result files are treated as replicates of the same sample.

Figure 300. Protein ratios when single quantification result files are treated as replicates of the same sample



The Proteome Discoverer application combines only protein quantification ratios from the same type of quantification—that is, either precursor-ion- or reporter-ion-based quantification—into replicate ratios. The names of the protein ratios must be the same to be combined into replicate ratios. For example, the ratios to combine into replicates must all be from reporter-ion-based quantification, and they must all be identically named (such as 115/114) in the result files to be combined. The application reports ratios from different types of quantification or ratios with different names as if they were treatments—that is, side by side on the protein level of the multiconsensus report.

Mixed Mode

You can also mix replicate and treatment mode. For example, you can load three result files from an iTRAQ 4plex experiment and two files from a SILAC experiment, and specify treating the quantification results as replicates. In this case, the Proteome Discoverer application tries to treat all defined protein quantification ratios as replicates, if possible. It reports everything else side by side at the protein level of the multiconsensus report. In this example, it calculates the combined averaged ratios from the three iTRAQ 4plex files and the two SILAC 2plex files, and reports the iTRAQ and SILAC ratios side by side, as shown in Figure 301 and Figure 302. In this way, the application can mimic complex experimental setups.

Figure 301. Opening a multiconsensus report from three iTRAQ and two SILAC files in replicate mode

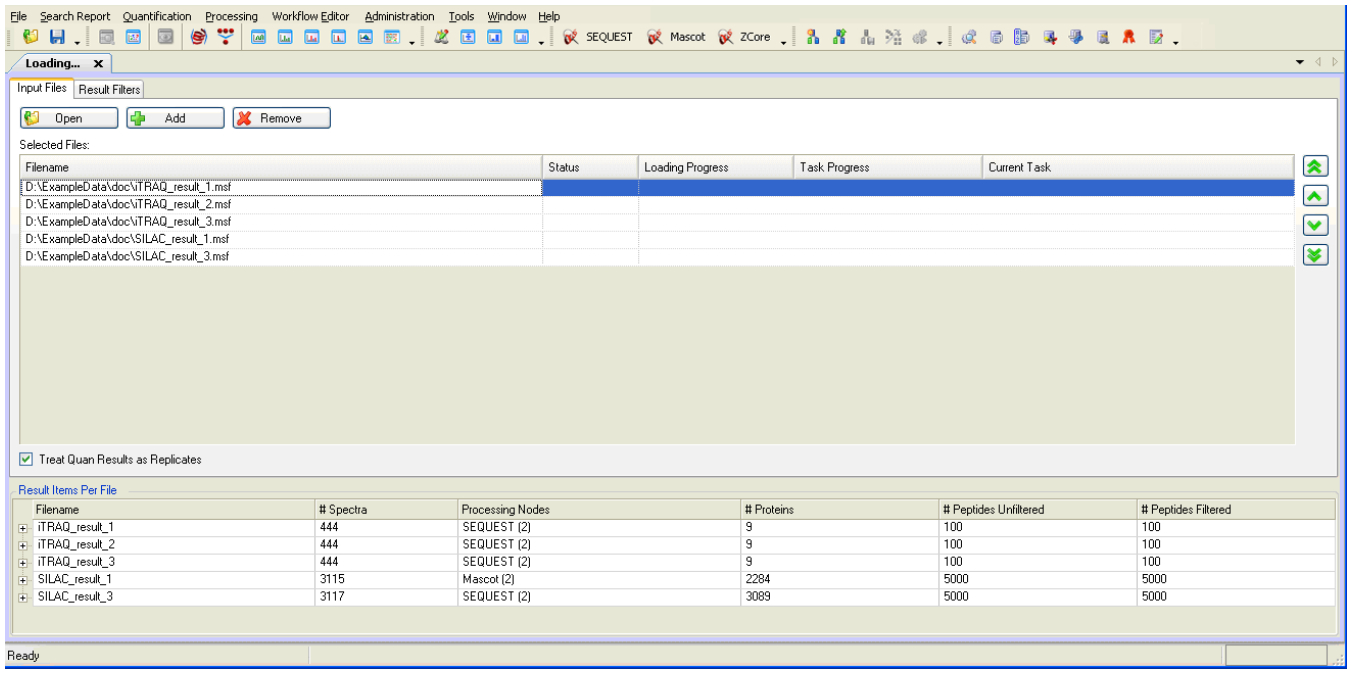
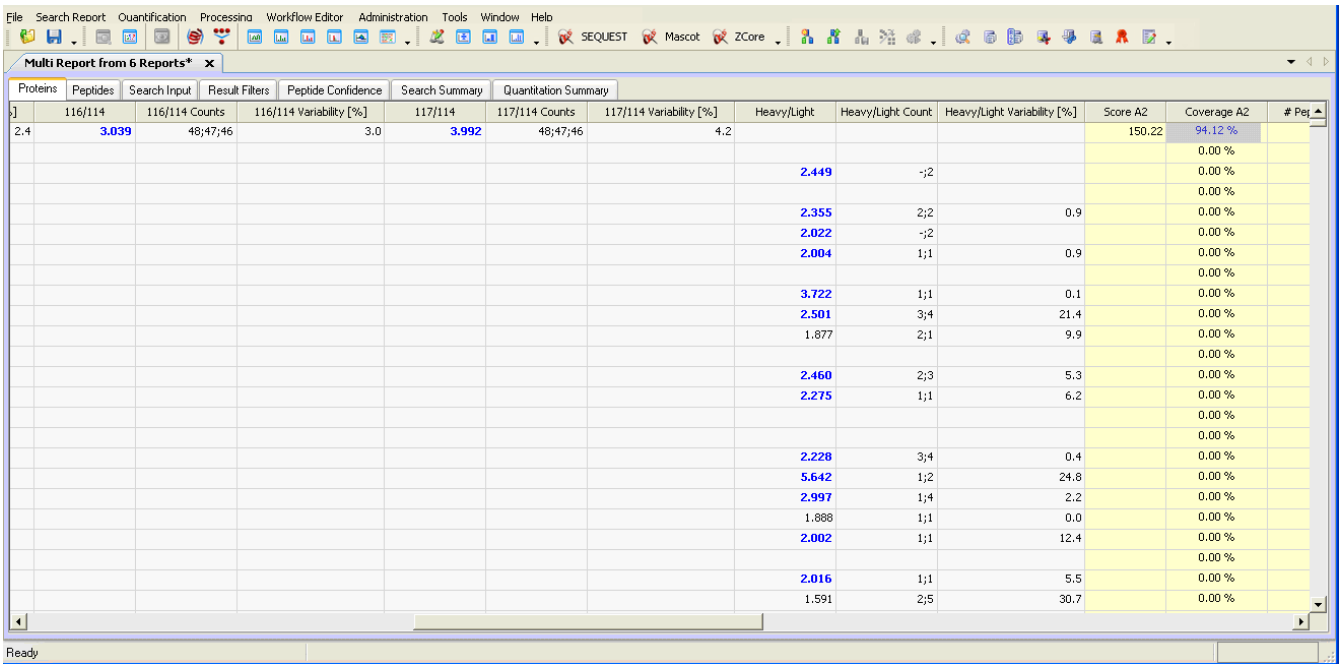


Figure 302 shows the opened multiconsensus report loaded in Figure 301. The combined ratios from the iTRAQ and the SILAC quantification are displayed side by side. In this example, the two searches are from different samples, and the two different types of quantification share no proteins.

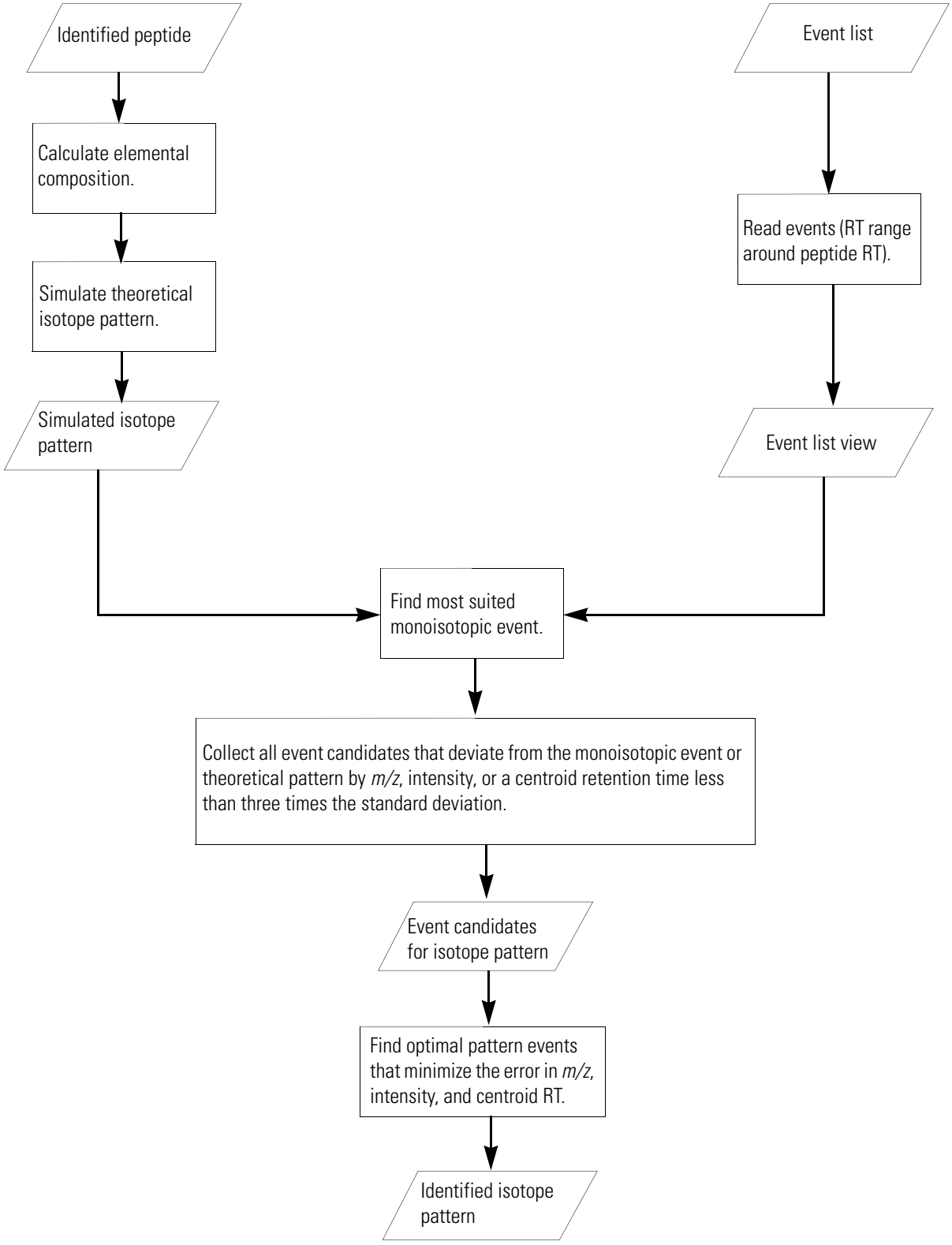
Figure 302. Opened multiconsensus report from three iTRAQ and two SILAC files in replicate mode



Identifying Isotope Patterns in Precursor Ion Quantification

The quantification spectra on the pages of the MSF report show the isotope pattern used for quantifying the peptides. The algorithm used in precursor ion quantification finds isotope patterns by identifying target components—that is, known elemental compositions from event lists. It identifies the peptides and searches in the event lists for the isotope patterns of these identified peptides. After peptide identification, the algorithm follows the steps shown in Figure 303 to identify the isotope patterns.

Figure 303. Identifying isotope patterns



Troubleshooting Quantification

See these sections for troubleshooting quantification results:

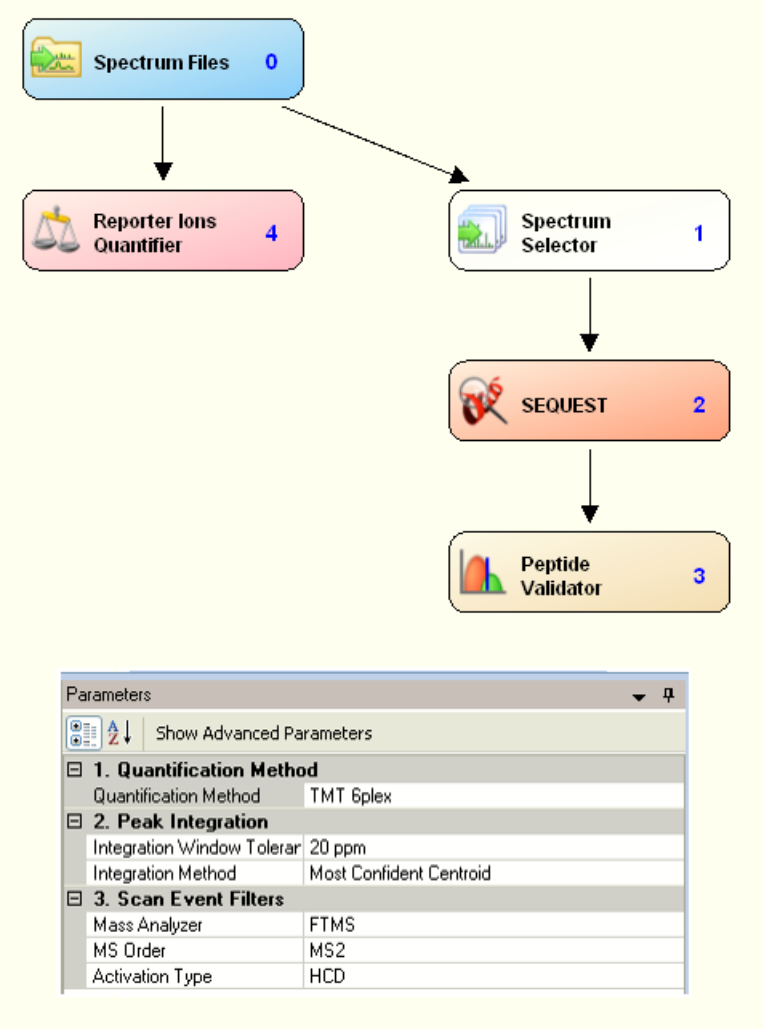
- Troubleshooting Reporter Ion Quantification
- Troubleshooting Precursor Ion Quantification

Troubleshooting Reporter Ion Quantification

If you obtain unexpected quantification results, verify that all settings of the nodes in your processing workflow are reasonable.

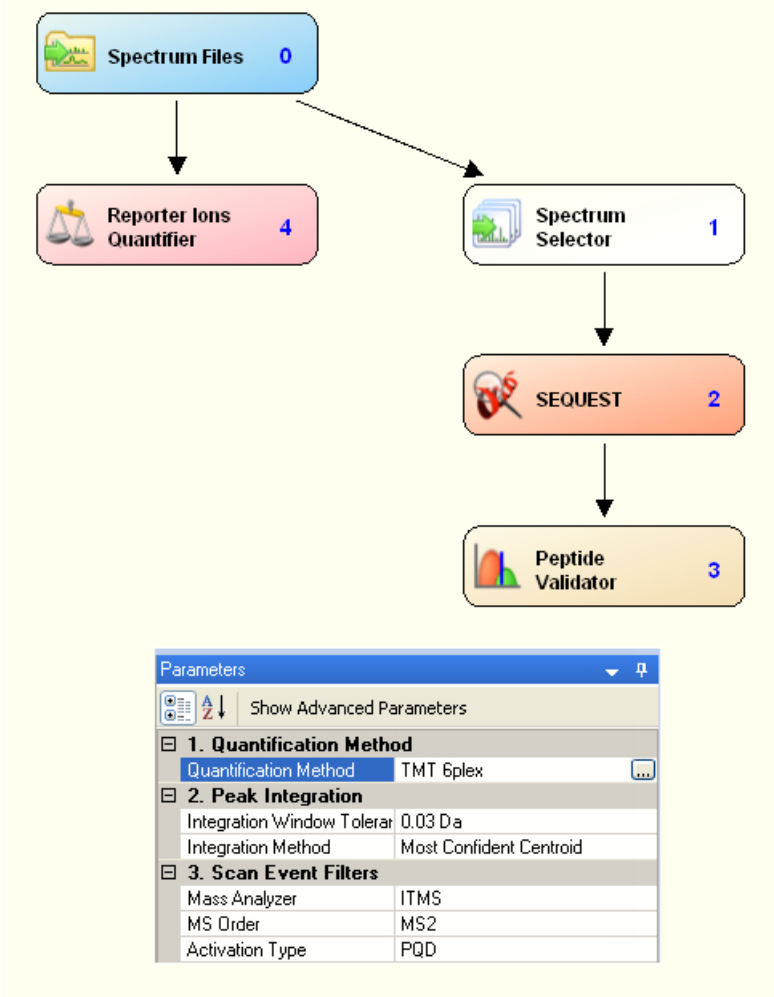
- Make sure that the Integration Window Tolerance parameter of the Reporter Ions Quantifier node fits the data that you are processing. The default is 20 ppm, which is too low if you are processing PQD data from an ion trap.
- Make sure that the settings of the Mass Analyzer, MS Order, and Activation Type parameters of the Reporter Ions Quantifier node are correct for the data that you are processing. Figure 304 shows the typical settings to use if you want to quantify HCD scans from the Orbitrap.

Figure 304. Typical settings for quantifying iTRAQ or TMT tags from HCD scans



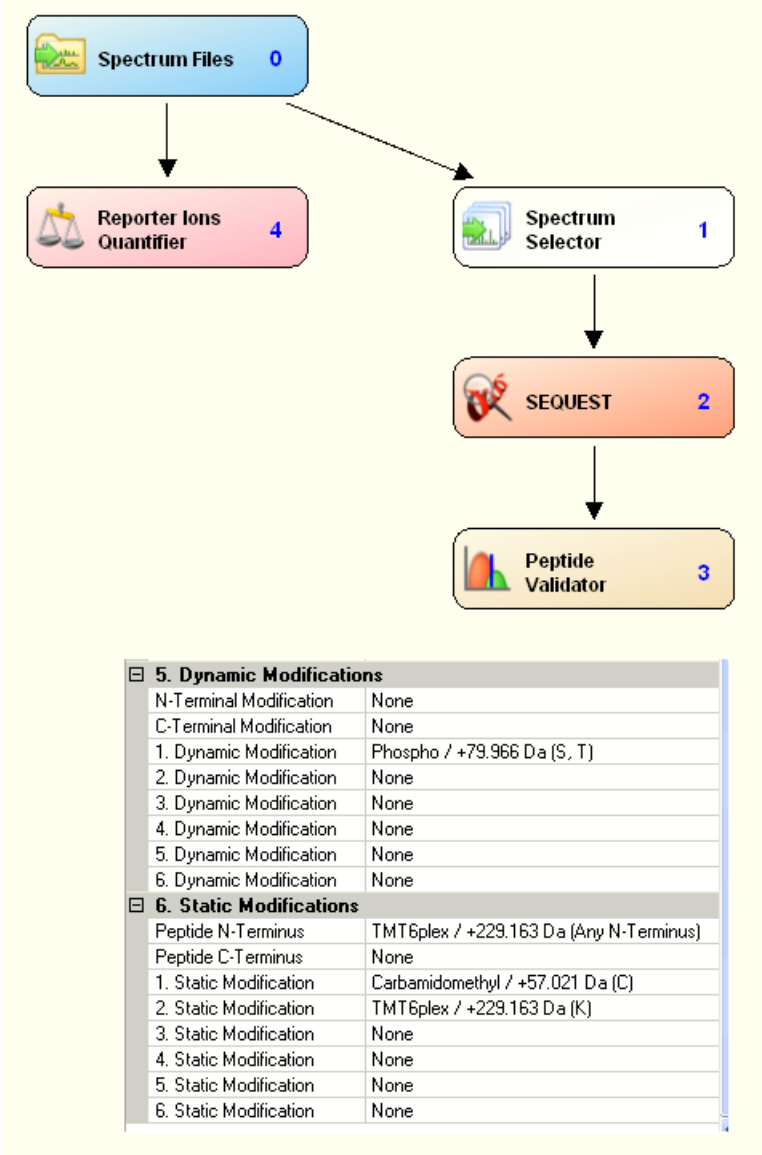
To quantify PQD scans from an ion trap, use the typical settings shown in Figure 305.

Figure 305. Typical settings for quantifying ITRAQ or TMT tags from the ion trap PQD scans



- Make sure that you have used the correct set of static and dynamic modifications for the search engine. For example, if you are searching TMT 6plex data with SEQUEST, check that your settings resemble Figure 306.

Figure 306. Modifications required for searching iTRAQ 4plex samples



5. Dynamic Modifications	
N-Terminal Modification	None
C-Terminal Modification	None
1. Dynamic Modification	Phospho / +79.966 Da (S, T)
2. Dynamic Modification	None
3. Dynamic Modification	None
4. Dynamic Modification	None
5. Dynamic Modification	None
6. Dynamic Modification	None
6. Static Modifications	
Peptide N-Terminus	TMT6plex / +229.163 Da (Any N-Terminus)
Peptide C-Terminus	None
1. Static Modification	Carbamidomethyl / +57.021 Da (C)
2. Static Modification	TMT6plex / +229.163 Da (K)
3. Static Modification	None
4. Static Modification	None
5. Static Modification	None
6. Static Modification	None

Troubleshooting Precursor Ion Quantification

If you obtain unexpected precursor ion quantification results, verify that all settings of your processing workflow are reasonable.

1. Check the dynamic modification parameters in the SEQUEST, Mascot, or ZCore search engines.

   These should match your isotope labeling sample.

2. Check the node parameters that you set before performing the quantification to see if they are appropriate for your sample.

   See "Performing Reporter Ion Quantification" on page 476 for more information.

3. Verify that your isotopic labeling is one of the following options in the protein ID/search node (either SEQUEST or Mascot):

   - SILAC 2plex (Arg10, Lys6): Uses arginine 10 and lysine 6.
   - SILAC 2plex (Arg10, Lys8): Uses arginine 10 and lysine 8.
   - SILAC 2plex (Ile6): Uses isoleucine 6.
   - SILAC 3plex (Arg6, Lys4|Arg10, Lys8): Uses arginine 10 and lysine 8 for "heavy" labels and arginine 6 and lysine 4 for "medium" labels.
   - SILAC 3plex (Arg6, Lys6|Arg10, Lys8): Uses arginine 10 and lysine 8 for "heavy" labels and arginine 6 and lysine 6 for "medium" labels.
   - Dimethylation 3plex: Chemically adds isotopically labeled dimethyl groups to the N-terminus and to the ε-amino group of lysine.
   - ^{18}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O.

   Note Low-mass accuracy cannot be used for precursor ion quantification or precursor ion area detection.

4. Check your tolerance window. If you get too many results, decrease the size of the window. For too few results, increase the size of the window.
5. Make sure you chose the right database.
6. Check the species listed to make sure the samples came from that species.
7. Verify that the activation type used is correct.
8. Verify that the instrument type in the Mascot search engine is correct.
9. Use only the ETD Spectrum Charger node for low-mass resolution ETD data.

Exporting Data

This chapter describes how to export your data and search results to other applications.

Contents

- Export Options
- Exporting Spectra
- Exporting Spectra Without a Peptide Match
- Exporting Annotated Spectra for Selected Proteins or Peptide Groups
- Exporting Search Results in ProtXML Format
- Exporting Search Results in PepXML Format
- Copying or Saving a View to an Image
- Exporting Protein References to a FASTA File
- Exporting Exclusion Mass Lists to Xcalibur
- Exporting Search Results to Excel
- Exporting Search Results to a Tab-Delimited TXT File

Export Options

The Proteome Discoverer application offers numerous options for exporting data and search results into other file formats, such as XML. You can do the following:

- Export spectra from the search results to MGF, MZDATA, DTA, and MZML file formats. You can export spectra even though they do not identify any peptides.
- Export search results to XML files.
- Export search results to tab-delimited TXT files.
- Export graphs to the Clipboard to save in EMF, PNG, GIF, JPG, TIFF, or BMP format.
- Export exclusion mass lists to the Xcalibur data system.
- Export data from the search results by pasting it into Excel files.

- Export protein references and sequences in the search results to a FASTA file.
- Export annotated spectra for selected peptides into a ZIP file that includes an HTML page with peptide information and links to spectrum images.

Exporting Spectra

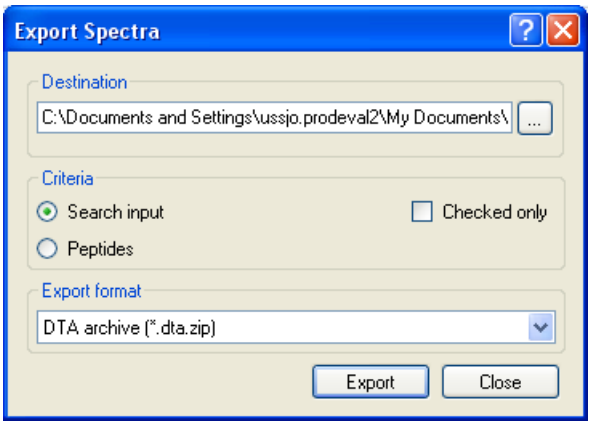
You can export spectra from the search results to MGF, MZDATA, DTA, and MZML file formats.

❖ To export spectra

1. Open your MSF file. See "Opening MSF Files" on page 215.
2. Choose **File > Export > Spectra**.

   The Export Spectra dialog box appears, as shown in Figure 307.

   Figure 307. Export Spectra dialog box



3. In the Destination box, type the path and name of the file, or click the browse button (…) to select the file where you want to save the exported spectra, and then click **Save** in the Save As dialog box.
4. In the Criteria area of the Export Spectra dialog box, select the type of data to export:
   - Search Input: Exports all spectra extracted from the raw file.
   - Peptides: Exports all spectra that identify a peptide.
   - Checked Only: Exports all spectra that identify peptides that you select.
5. In the Export Format list, select the format of the file to export the spectra to:
   - DTA Archive file: A file containing MS^n data for single or grouped scans.
   - Mascot Generic Format (MGF) file: A mass spectral file produced during Mascot analysis. It contains a list of precursor ions, their fragments, and the masses of the fragments.

- MZDATA file: Common data format files developed by the Human Proteome Organization (HUPO) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line.
- MZML files are a combination of MZDATA and MZXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the Seattle Proteome Center at the Institute for Systems Biology (ISB). The Proteome Discoverer application supports version 1.1.0 of the MZML format.

6. Click **Export**.

   When the exportation is complete, you can open the exported file.

   –or–

   To close the Export Spectra dialog box without exporting any spectra or applying any of the selected options, click **Close**.

Export Spectra Dialog Box Parameters

Table 95 lists the parameters in the Export Spectra dialog box.

Table 95. Export Spectra dialog box (Sheet 1 of 2) parameters

Command or option	Description
Destination	Specifies the path and name of the file where you want to save the exported spectra.
Criteria	Selects the type of spectra to export.
Search input	Exports all spectra extracted from the raw file.
Peptides	Exports all spectra that identify a peptide.
Checked only	Exports all spectra that identify the peptides or Search Input that you select.

Table 95. Export Spectra dialog box (Sheet 2 of 2) parameters

Command or option	Description
Export format	Specifies the format of the file to save the exported search results in. The available formats are the following: • DTA Archive: A file containing MS^n data for single or grouped scans. • Mascot Generic Format (MGF): A mass spectral file produced during Mascot analysis. It contains a list of precursor ions, their fragments, and the masses of the fragments. • MZDATA: A common data format file developed by the Human Proteome Organization (HUPO) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line. • MZML files are a combination of MZDATA and MZXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the Seattle Proteome Center at the Institute for Systems Biology (ISB). The Proteome Discoverer application supports version 1.1.0 of the MZML format.
Export	Saves the selected search results in the specified destination file.
Close	Closes the dialog box without exporting any spectra or applying any of the selected options.

Exporting Spectra Without a Peptide Match

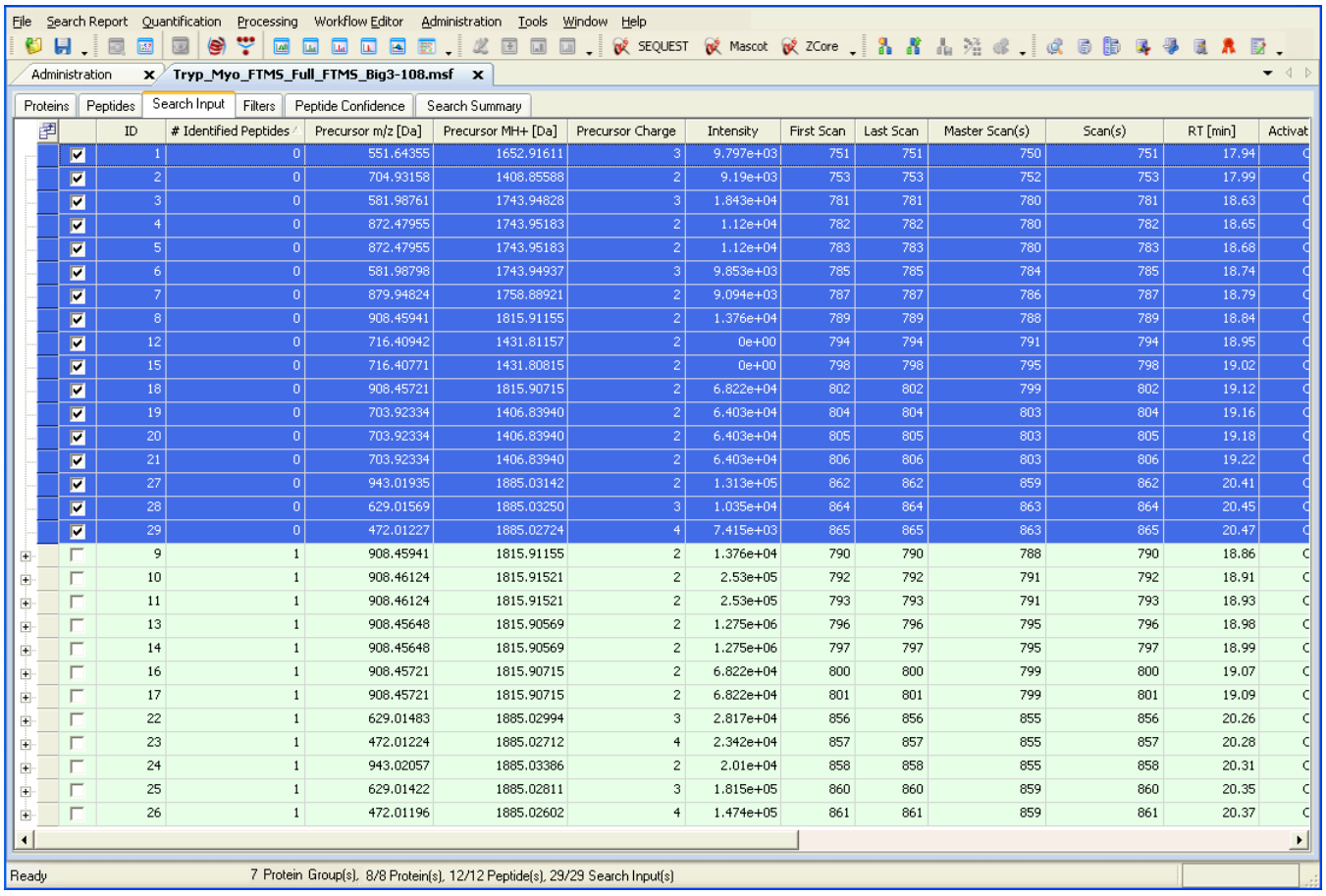
You can export spectra even though they do not identify any peptides.

❖ To export spectra without a peptide match

1. On the Search Input page, sort the list by clicking the # Identified Peptides column header.
2. Select all rows with a reported number of zero peptide matches by using one of the following methods:
   - Select the first row by clicking the gray box at the beginning of the row and then dragging the cursor up or down to select additional rows.
   - Select the first row by clicking the gray box at the beginning of the row. Release the mouse button. Holding down the SHIFT key, select the last row. You can select or clear single rows by holding down the CTRL key.
3. Set the check mark for all selected rows by pressing the SPACEBAR.

   The result should resemble the example in Figure 308.

Figure 308. Setting check marks on selected items on the Search Input page



ID	# Identified Peptides	Precursor m/z [Da]	Precursor MH+ [Da]	Precursor Charge	Intensity	First Scan	Last Scan	Master Scan(s)	Scan(s)	RT [min]	Activat
1	0	551.64355	1652.91611	3	9.797e+03	751	751	750	751	17.94	C
2	0	704.93158	1408.85588	2	9.19e+03	753	753	752	753	17.99	C
3	0	581.98761	1743.94828	3	1.843e+04	781	781	780	781	18.63	C
4	0	872.47955	1743.95183	2	1.12e+04	782	782	780	782	18.65	C
5	0	872.47955	1743.95183	2	1.12e+04	783	783	780	783	18.68	C
6	0	581.98798	1743.94937	3	9.853e+03	785	785	784	785	18.74	C
7	0	879.94824	1758.88921	2	9.094e+03	787	787	786	787	18.79	C
8	0	908.45941	1815.91155	2	1.376e+04	789	789	788	789	18.84	C
12	0	716.40942	1431.81157	2	0e+00	794	794	791	794	18.95	C
15	0	716.40771	1431.80815	2	0e+00	798	798	795	798	19.02	C
18	0	908.45721	1815.90715	2	6.822e+04	802	802	799	802	19.12	C
19	0	703.92334	1406.83940	2	6.403e+04	804	804	803	804	19.16	C
20	0	703.92334	1406.83940	2	6.403e+04	805	805	803	805	19.18	C
21	0	703.92334	1406.83940	2	6.403e+04	806	806	803	806	19.22	C
27	0	943.01935	1885.03142	2	1.313e+05	862	862	859	862	20.41	C
28	0	629.01569	1885.03250	3	1.035e+04	864	864	863	864	20.45	C
29	0	472.01227	1885.02724	4	7.415e+03	865	865	863	865	20.47	C
9	1	908.45941	1815.91155	2	1.376e+04	790	790	788	790	18.86	C
10	1	908.46124	1815.91521	2	2.53e+05	792	792	791	792	18.91	C
11	1	908.46124	1815.91521	2	2.53e+05	793	793	791	793	18.93	C
13	1	908.45648	1815.90569	2	1.275e+06	796	796	795	796	18.98	C
14	1	908.45648	1815.90569	2	1.275e+06	797	797	795	797	18.99	C
16	1	908.45721	1815.90715	2	6.822e+04	800	800	799	800	19.07	C
17	1	908.45721	1815.90715	2	6.822e+04	801	801	799	801	19.09	C
22	1	629.01483	1885.02994	3	2.817e+04	856	856	855	856	20.26	C
23	1	472.01224	1885.02712	4	2.342e+04	857	857	855	857	20.28	C
24	1	943.02057	1885.03386	2	2.01e+04	858	858	855	858	20.31	C
25	1	629.01422	1885.02811	3	1.815e+05	860	860	859	860	20.35	C
26	1	472.01196	1885.02602	4	1.474e+05	861	861	859	861	20.37	C

4. Choose **File > Export > Spectra**.

   The Export Spectra dialog box opens, as shown in Figure 307.

5. Select the export format from the Export Format list.
6. In the Criteria area, select the **Search Input** option and the **Checked Only** check box.
7. In the Destination box, type the file name and path where you want to save the exported spectra, or click the browse button (…), browse for the file in the Save As dialog box, and click **Save**.
8. In the Export Spectra dialog box, click **Export**.

Exporting Annotated Spectra for Selected Proteins or Peptide Groups

You can export peptide spectra annotated with observed masses and fragment assignments in an HTML file. You can export all peptide spectra, selected peptide spectra, the peptides of selected proteins, or the peptides of selected peptide groups.

A peptide group contains all the reported peptides that have the same sequence, exclusive of modifications. For example, a peptide with the ADLF sequence that has no modifications belongs to the same peptide group as a peptide with the ADLF sequence that has a modified leucine and a peptide with the ADLF sequence that has a terminal modification.

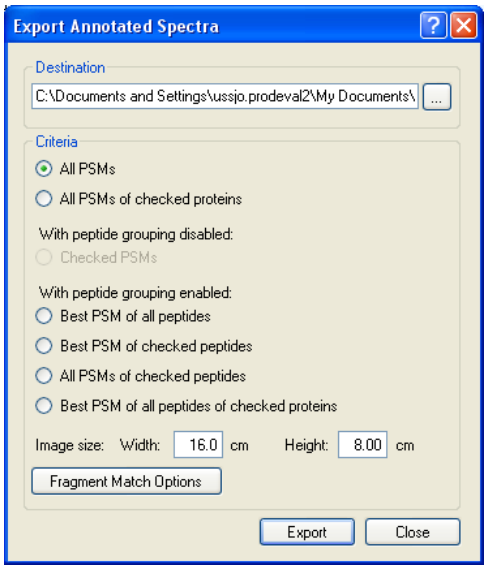
The Proteome Discoverer application exports only the peptides that are not filtered out.

❖ To export annotated spectra for peptides

1. Open your search results in the MSF file.
2. Choose **File > Export > Annotated Spectra**.

   The Export Annotated Spectra dialog box opens, as shown in Figure 309.

Figure 309. Export Annotated Spectra dialog box



3. In the Destination box, do one of the following:

   - Type the path and name of the ZIP file in which to store the image and the TXT file for the annotated peptides.

     –or–

   - Click the browse button, and in the Save As dialog box, browse to the ZIP file or type the name of the ZIP file in which to store the image and .the txt file for the annotated peptides, and then click **Save**.

4. In the Criteria area, specify the type of peptides to export:

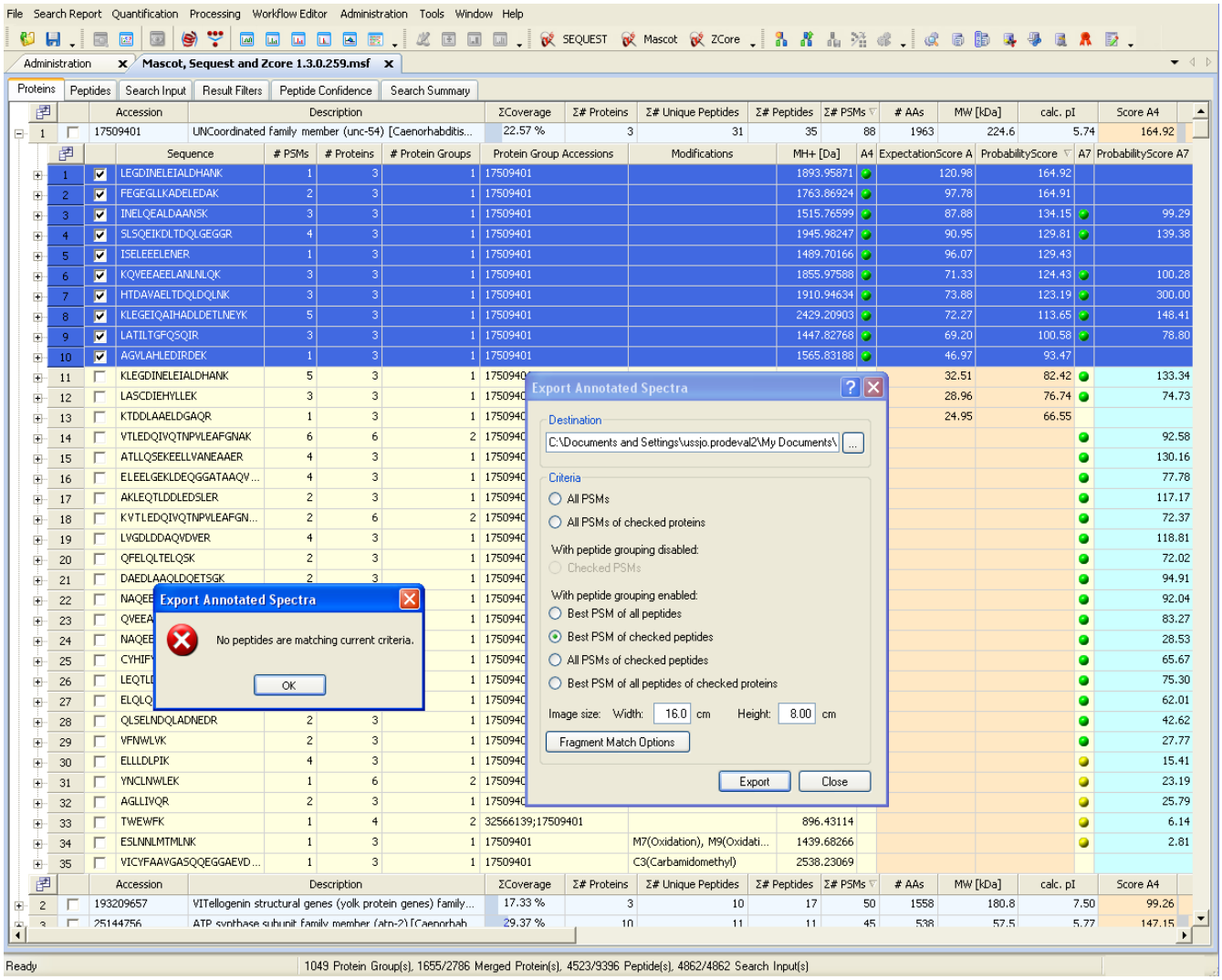
   Note To select the peptides or proteins to export, place a check mark to the left of the names of the target proteins, protein groups, peptides, or peptide groups on the Proteins or Peptides pages.

   - All PSMs: Exports the annotated spectra of all peptide spectrum matches in the report that are not filtered out.
   - All PSMs of Checked Proteins: Exports the annotated spectra of the peptide spectrum matches belonging to selected proteins.
   - With Peptide Grouping Disabled: Makes the following options available only when peptide grouping is not enabled:
     – Checked PSMs: Exports the annotated spectra of the peptide spectrum matches belonging to selected peptides.

- With Peptide Grouping Enabled: Makes the following options available only when peptide grouping is enabled:
  - Best PSM of All Peptides: Exports the annotated spectra of the best peptide spectrum match of every peptide.
  - Best PSM of Checked Peptides: Exports the annotated spectra of the best peptide spectrum match for every selected peptide.
  - All PSMs of Checked Peptides: Exports the annotated spectra of all peptide spectrum matches of every selected peptide.
  - Best PSM of All Peptides of Checked Proteins: Exports the best peptide spectrum match for every peptide of all selected proteins. The HTML index page contained in the exported zip archive lists the proteins with their annotations and descriptions followed by their PSMs.

  You must click the Peptides tab to activate the Peptides page before you select any of the With Peptide Grouping Enabled options. If you do not, the error message shown in Figure 310 appears.

Figure 310. Error message issued when Peptides page is not selected



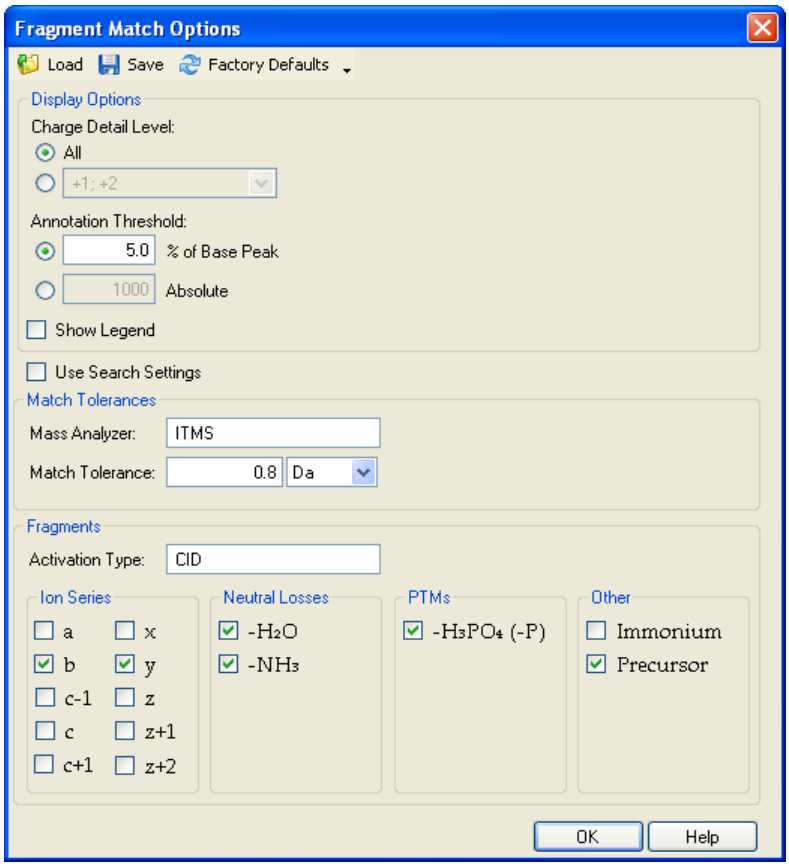
5. In the Image Size boxes, specify the width and height of the spectra.

   The default width is 16.00 centimeters, and the default height is 8.00 centimeters.

6. Click **Fragment Match Options**.

   The Fragment Match Options dialog box opens, as shown in Figure 311.

Figure 311. Fragment Match Options dialog box



For information on the parameters on this dialog box, see "Options Dialog Box Parameters" on page 56.

a. To load the most recently saved settings, click **Load** (Load).

b. Select a state from the Charge Detail Level list. You can choose from the following settings:

- (Default) All: Displays all fragment ions, regardless of charge.
- +1: Displays singly charged fragment ions.
- +2: Displays doubly charged fragment ions.
- +3: Displays triply charged fragment ions.
- +4 …: Displays additional charged fragment ions if available.

c. In the Annotation Threshold area, specify a peak intensity limit.

   Peaks below this limit are not annotated. This limit prevents the addition of too many labels to a spectrum, especially if the low-intensity peaks are noise peaks. You can select either of the following methods:

   - % of Base Peak (Default): Specifies an intensity threshold as a percentage of the base peak in the spectrum. The base peak is the largest peak in a spectrum.
   - Absolute: Specifies an absolute intensity threshold.

d. When you want colored lines denoting a, b, and y ions to appear at the bottom of each spectrum image, select the **Show Legend** check box.

e. To use the match tolerances and ion series from the search, select the **Use Search Settings** check box.

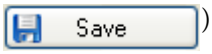
f. In the Mass Analyzer list, select the type of mass analyzer that was used to obtain the spectrum.

g. Specify the mass tolerance. The Proteome Discoverer application annotates the spectrum with ions that it assigns within this mass tolerance. You can use daltons (Da) or parts per million (PPM).

h. In the Fragments area, select the type of fragment ions to include in the annotated exported spectra:

   i. Select the activation type from the Activation Type list.

   ii. Select the check boxes corresponding to the fragment ion type.

i. To save the settings as the default settings, click **Save** (

).

j. Click **OK** to apply the Fragment Match Options settings and close the Fragment Match Options dialog box.

7. In the Export Annotated Spectra dialog box, click **Export**.

   –or–

   To close the Export Annotated Spectra dialog box without exporting any spectra or applying any of the selected options, click **Close**.

If you click Export, a meter in the lower left corner of the Export Annotated Spectra dialog box indicates the progress of the exportation.

When exporting the file is complete, you can open it.

Tip Use the 7zip or WinZip decompression program. If you compress an output HTML file and later attempt to open it with another decompression program by right-clicking the file and choosing Open With > Compressed (zipped) Folders, the extraction might take a very long time and damage the contents of the file.

The exported HTML file consists of two parts:

- A peptides table containing a list of the found peptides. It includes links to the annotated spectrum and peak list for each peptide, as well as the peptide confidence.
- A search summary containing the same information as the Search Summary page of an open report. Figure 316 shows a sample excerpt from this summary. For detailed information about the Search Summary page, see "Using the Search Summary Page" on page 349.

Peptides Table in the Output HTML File

Figure 312 shows a sample excerpt from a peptides table in an output HTML file showing the annotated spectra for all peptides.

Figure 312. Peptides table in the HTML file showing annotated spectra for all peptides

Annotated spectra of helaRVD1_1_1_1 40-60 min.msf

go to Peptides
go to Search Summary

Peptides

Confidence	Sequence	Activation Type	Modifications	IonScore	ΔScore	Rank	Search Engine	Charge	Precursor m/z [Da]	ΔM [ppm]	First Scan	Last Scan	Annotated Spectrum	Peak List
Low	DYGNSPLHR	CID		20.44	1.00	1	Mascot (2)	2	529.75592	2.98	5573	5573	Image	Peak List
Low	NSDFIk	CID	K6(Lys6)	10.29	1.00	1	Mascot (2)	2	365.19916	5.07	5574	5574	Image	Peak List
Low	VAEWR	CID		15.99	0.31	1	Mascot (2)	2	330.67780	2.87	5576	5576	Image	Peak List
Low	LGEWR	CID		10.71		2	Mascot (2)	2	330.67780	2.87	5576	5576	Image	Peak List
Low	SQEGSQFVk	CID	K9(Lys8)	15.23	1.00	1	Mascot (2)	2	509.25922	1.98	5578	5578	Image	Peak List
Low	SckQTQMKNK	CID	C2(Carbamidomethyl), K3(Lys6)	11.96	0.08	1	Mascot (2)	2	629.82495	6.98	5579	5579	Image	Peak List
Low	mANVELSSLAr	CID	N-Term(Acetyl), M1(Oxidation), R11(Arg10)	10.98		2	Mascot (2)	2	629.82495	7.17	5579	5579	Image	Peak List
Low	GTPGPPPAHGAALQFHPR	CID		12.97	0.15	1	Mascot (2)	3	586.64648	3.07	5580	5580	Image	Peak List
Low	mVmAAKkGPGPGGGVSGGk	CID	N-Term(Acetyl), M3(Oxidation), K7(Lys6), K19(Lys8)	11.03		2	Mascot (2)	3	586.64648	1.37	5580	5580	Image	Peak List
Low	mVMAAKkGPGPGGGVSGGk	CID	N-Term(Acetyl), M1(Oxidation), K7(Lys6), K19(Lys8)	10.69		3	Mascot (2)	3	586.64648	1.37	5580	5580	Image	Peak List
High	FDDGAGGDNEVQr	CID	R13(Arg10)	66.20	1.00	1	Mascot (2)	2	695.30121	3.30	5581	5581	Image	Peak List
High	LESGmQNMSIHTK	CID	M5(Oxidation)	53.32	0.40	1	Mascot (2)	2	746.35223	2.75	5582	5582	Image	Peak List
Medium	LESGMQNmSIHTK	CID	M8(Oxidation)	32.30		2	Mascot (2)	2	746.35223	2.75	5582	5582	Image	Peak List
Low	kELQEEEREk	CID	K1(Lys6), K10(Lys8)	28.38	0.18	1	Mascot (2)	2	666.35718	6.16	5583	5583	Image	Peak List
Low	EkLQEEmLQr	CID	K2(Lys6), M7(Oxidation), R10(Arg8)	23.10		2	Mascot (2)	2	666.35718	3.28	5583	5583	Image	Peak List

Link to annotated spectrum

Link to peak list

Figure 313 shows a sample excerpt of an HTML file showing the annotated spectra for only selected peptides.

Figure 313. Peptides table in the HTML file showing annotated spectra for selected peptides

Annotated spectra of helaRVD1_1_1_1 40-60 min.msf

go to Peptides
go to Search Summary

Peptides

Confidence	Sequence	Activation Type	Modifications	IonScore	ΔScore	Rank	Search Engine	Charge	Precursor m/z [Da]	ΔM [ppm]	First Scan	Last Scan	Annotated Spectrum	Peak List
High	GESPVDYDGGr	CID	R.11(Arg10)	44.33	0.66	1	Mascot (2)	2	581.25793	3.37	6206	6206	Image	Peak List
High	AQYEDIANr	CID	R.9(Arg10)	44.39	0.73	1	Mascot (2)	2	545.26569	3.83	6252	6252	Image	Peak List
Medium	LSSPATLNSR	CID		44.73	0.73	1	Mascot (2)	2	523.28741	3.69	6383	6383	Image	Peak List
High	NDQcYDDIR	CID	C.4(Carbamidomethyl)	44.48	0.61	1	Mascot (2)	2	599.74634	5.03	6400	6400	Image	Peak List
High	LMIEmDGTENK	CID	M.5(Oxidation)	44.58	0.51	1	Mascot (2)	2	648.79657	3.68	6899	6899	Image	Peak List
High	IDEPLEGSEDr	CID	R.11(Arg10)	44.68	1.00	1	Mascot (2)	2	635.29736	3.24	6917	6917	Image	Peak List
Medium	GVTIKPTVDDD	CID		44.41	1.00	1	Mascot (2)	2	580.29779	3.56	7067	7067	Image	Peak List
Medium	LEQDEYALR	CID		44.82	0.76	1	Mascot (2)	2	568.78430	2.69	7587	7587	Image	Peak List
Medium	QmSLLLR	CID	M.2(Oxidation)	44.76	0.33	1	Mascot (2)	2	438.75412	4.30	8354	8354	Image	Peak List
Medium	YYVTIIDAPGHR	CID		44.46	0.73	1	Mascot (2)	3	468.91599	4.51	9571	9571	Image	Peak List
Medium	GIYAYGFEkPSAIQQr	CID	K.9(Lys8), R.16(Arg10)	44.35	1.00	1	Mascot (2)	3	615.99481	4.63	9736	9736	Image	Peak List
Medium	LAVEAVLr	CID	R.8(Arg10)	44.29		2	Mascot (2)	2	440.77994	4.03	9777	9777	Image	Peak List
Medium	AFIITGQTYTr	CID	R.11(Arg10)	44.47	0.73	1	Mascot (2)	2	640.85034	4.75	9809	9809	Image	Peak List
Medium	VLISDSLDPccR	CID	C.10(Carbamidomethyl), C.11 (Carbamidomethyl)	44.48	1.00	1	Mascot (2)	2	717.84259	4.37	10013	10013	Image	Peak List
Medium	LLIVSTTPYSEK	CID		44.77	0.69	1	Mascot (2)	2	675.88159	3.22	10146	10146	Image	Peak List
Medium	DQLLLGPTYATPK	CID		44.30	0.70	1	Mascot (2)	2	708.89319	4.05	10690	10690	Image	Peak List

Figure 314 shows a sample excerpt of an HTML file showing annotated spectra for the peptides belonging to selected proteins

Figure 314. Peptides table in the HTML file showing annotated spectra for peptides belonging to selected proteins

Annotated spectra of helaRVD1_1_1_1 40-60 min.msf

go to Peptides
go to Search Summary

Peptides

Accession - Description														
Confidence	Sequence	Activation Type	Modifications	IonScore	ΔScore	Rank	Search Engine	Charge	Precursor m/z [Da]	ΔM [ppm]	First Scan	Last Scan	Annotated Spectrum	Peak List
IPI:IPI00440493.2 - ATP synthase subunit alpha, mitochondrial														
Low	LTELLK	CID		29.03	0.31	2	Mascot (2)	2	358.73251	3.42	7817	7817	Image	Peak List
Low	APGIIPR	CID		23.74	0.33	1	Mascot (2)	2	362.23062	3.78	7011	7011	Image	Peak List
Low	LTELLk	CID	K.6(Lys8)	16.15	0.31	2	Mascot (2)	2	362.73969	3.58	7832	7832	Image	Peak List
Low	APGIIPr	CID	R.7(Arg10)	33.11	0.18	1	Mascot (2)	2	367.23474	3.69	7030	7030	Image	Peak List
Low	ELIIGDR	CID		16.60	0.41	1	Mascot (2)	2	408.23563	2.19	7896	7896	Image	Peak List
Low	ELIIGDr	CID	R.7(Arg6)	19.69	0.05	1	Mascot (2)	2	411.24609	3.15	7911	7911	Image	Peak List
Low	ELIIGDr	CID	R.7(Arg10)	25.15	0.40	1	Mascot (2)	2	413.24033	3.53	7920	7920	Image	Peak List
Medium	STVAQLVK	CID		37.72	0.53	1	Mascot (2)	2	423.25986	3.91	7246	7246	Image	Peak List
Low	STVAQLVk	CID	K.8(Lys8)	33.77	0.59	1	Mascot (2)	2	427.26666	3.18	7244	7244	Image	Peak List
Low	QMSLLLR	CID		29.09	0.34	1	Mascot (2)	2	430.75659	4.21	10411	10411	Image	Peak List
Low	QMSLLLr	CID	R.7(Arg10)	32.86	0.12	1	Mascot (2)	2	435.76041	3.43	10390	10390	Image	Peak List
Medium	QmSLLLR	CID	M.2(Oxidation)	44.76	0.33	1	Mascot (2)	2	438.75412	4.30	8354	8354	Image	Peak List
Low	QmSLLLr	CID	M.2(Oxidation), R.7 (Arg6)	31.95		2	Mascot (2)	2	441.76407	4.01	8405	8405	Image	Peak List
Low	LELAQYR	CID		23.41	0.26	1	Mascot (2)	2	446.74969	3.80	8166	8166	Image	Peak List
Low	LELAQYr	CID	R.7(Arg6)	22.19	0.23	1	Mascot (2)	2	449.75983	3.92	8156	8156	Image	Peak List

Figure 315 shows a sample excerpt of an HTML file showing the annotated spectra of the peptides of selected peptide groups

Figure 315. Peptides table in the HTML file showing annotated spectra of the peptides of selected peptides groups

Annotated spectra of helaRVD1_1_1_1 40-60 min.msf

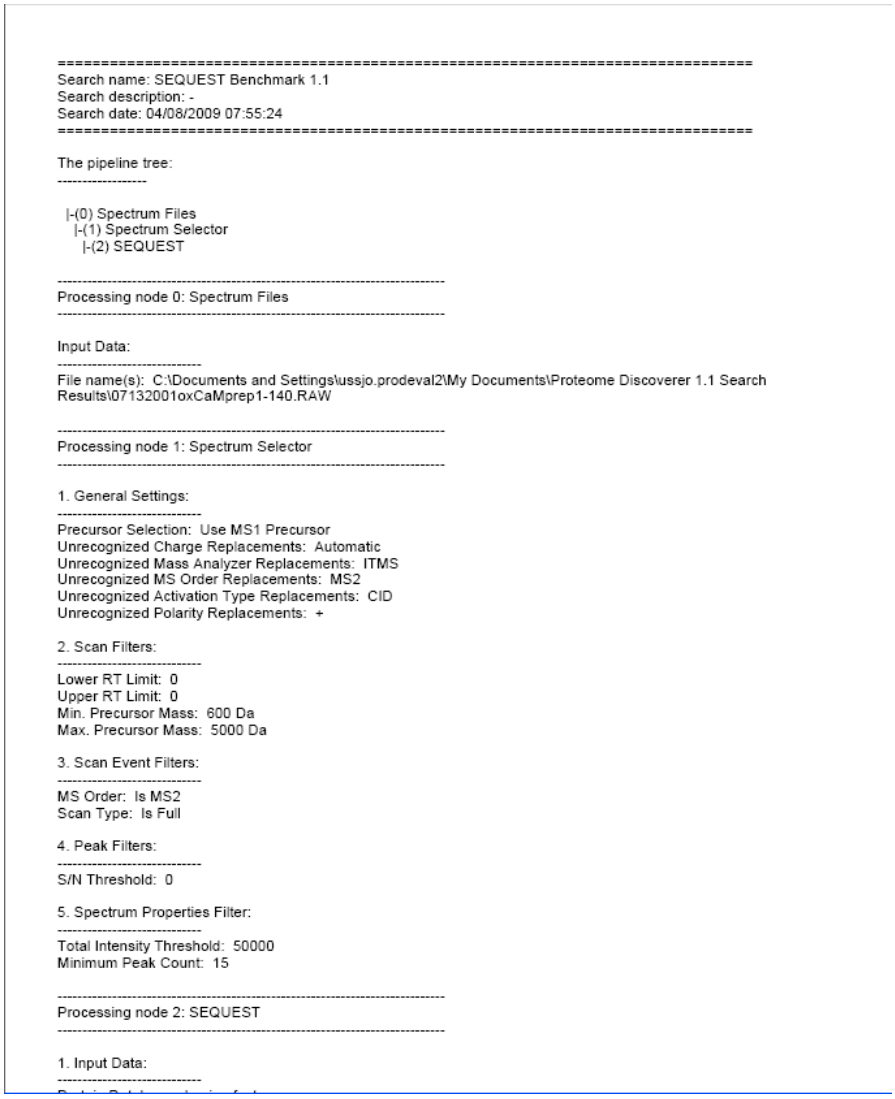
go to Peptides
go to Search Summary

Peptides

Sequence															
	Confidence	Sequence	Activation Type	Modifications	IonScore	ΔScore	Rank	Search Engine	Charge	Precursor m/z [Da]	ΔM [ppm]	First Scan	Last Scan	Annotated Spectrum	Peak List
ALVVNER															
	Low	ALVVNER	CID		21.74		3	Mascot (2)	2	400.73602	2.81	5763	5763	Image	Peak List
	Low	ALVVNEr	CID	R7(Arg6)	25.32		2	Mascot (2)	2	403.74588	2.27	5793	5793	Image	Peak List
	Medium	ALVVNEr	CID	R7(Arg10)	37.44	0.00	1	Mascot (2)	2	405.74033	3.19	5800	5800	Image	Peak List
ALKGGLSR															
	Medium	ALKGGLSr	CID	R8(Arg10)	37.65	0.29	1	Mascot (2)	2	406.25696	5.19	9444	9444	Image	Peak List
STVAQLVK															
	Medium	STVAQLVK	CID		37.72	0.53	1	Mascot (2)	2	423.25986	3.91	7246	7246	Image	Peak List
	Low	STVAQLVk	CID	K8(Lys8)	33.77	0.59	1	Mascot (2)	2	427.26666	3.18	7244	7244	Image	Peak List
YYGAQTVR															
	Medium	YYGAQTVr	CID	R8(Arg10)	37.36	0.51	1	Mascot (2)	2	484.24908	3.88	5690	5690	Image	Peak List
LPLQDVVK															
	Low	LPLQDVYK	CID		26.70	0.52	1	Mascot (2)	2	488.28033	2.46	9122	9122	Image	Peak List
	Medium	LPLQDVYK	CID		37.35	0.46	1	Mascot (2)	2	488.28088	3.59	9297	9297	Image	Peak List
	Low	LPLQDVYk	CID	K8(Lys8)	21.96	0.23	1	Mascot (2)	2	492.28720	1.97	9113	9113	Image	Peak List
	Low	LPLQDVYk	CID	K8(Lys8)	34.13	0.59	1	Mascot (2)	2	492.28851	4.64	9275	9275	Image	Peak List
ATAEVLNIGK															
	Medium	aTAEVLNIGK	CID	N-Term(Acetyl)	38.02	1.00	1	Mascot (2)	2	529.29382	-8.02	11204	11204	Image	Peak List
AVDDGVNTFK															
	Medium	AVDDGVNTFK	CID		37.92	0.71	1	Mascot (2)	2	533.26599	3.32	7127	7127	Image	Peak List
	Low	AVDDGVNTFk	CID	K10(Lys8)	25.77	1.00	1	Mascot (2)	2	537.27356	4.17	7178	7178	Image	Peak List
FETLQAQAGK															
	Medium	FETLQAQAGK	CID		37.27	0.59	1	Mascot (2)	2	546.78979	3.53	6702	6702	Image	Peak List

Figure 316 Shows the search summary in the HTML file.

Figure 316. Search summary in the HTML file



```
================================================================================
Search name: SEQUEST Benchmark 1.1
Search description: -
Search date: 04/08/2009 07:55:24
================================================================================

The pipeline tree:
------------------

  |-(0) Spectrum Files
    |-(1) Spectrum Selector
      |-(2) SEQUEST

--------------------------------------------------------------------------------
Processing node 0: Spectrum Files
--------------------------------------------------------------------------------

Input Data:
-----------------------------
File name(s):  C:\Documents and Settings\ussjo.prodeval2\My Documents\Proteome Discoverer 1.1 Search
Results\07132001oxCaMprep1-140.RAW

--------------------------------------------------------------------------------
Processing node 1: Spectrum Selector
--------------------------------------------------------------------------------

1. General Settings:
-----------------------------
Precursor Selection:  Use MS1 Precursor
Unrecognized Charge Replacements:  Automatic
Unrecognized Mass Analyzer Replacements:  ITMS
Unrecognized MS Order Replacements:  MS2
Unrecognized Activation Type Replacements:  CID
Unrecognized Polarity Replacements:  +

2. Scan Filters:
-----------------------------
Lower RT Limit:  0
Upper RT Limit:  0
Min. Precursor Mass:  600 Da
Max. Precursor Mass:  5000 Da

3. Scan Event Filters:
-----------------------------
MS Order:  Is MS2
Scan Type:  Is Full

4. Peak Filters:
-----------------------------
S/N Threshold:  0

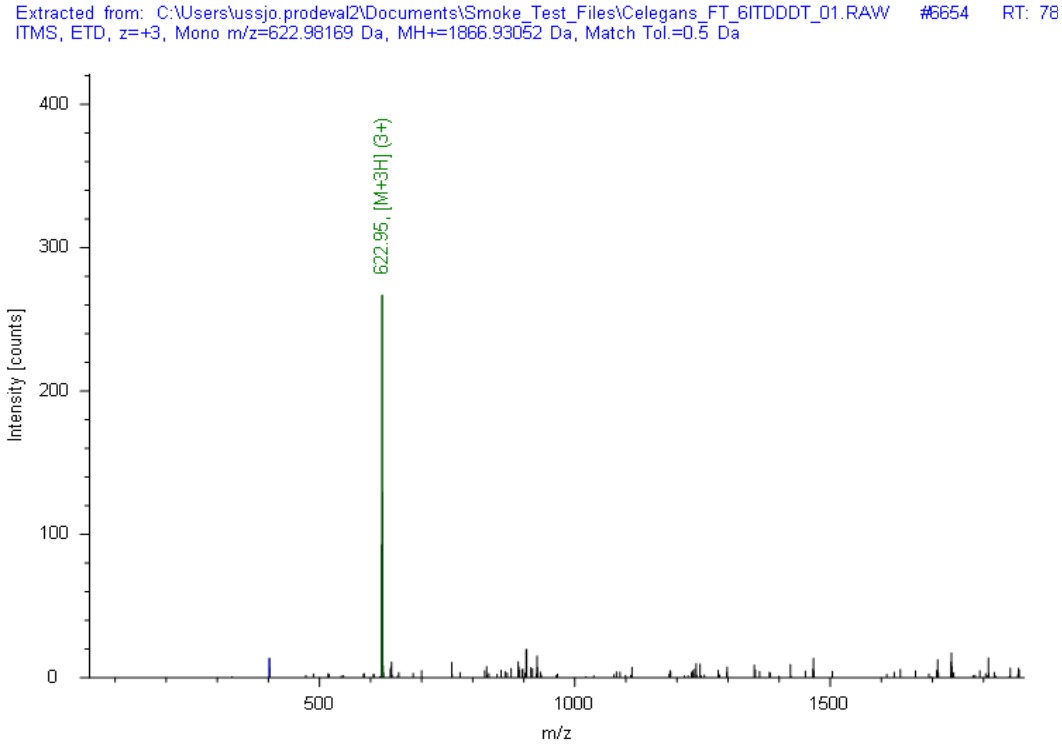
5. Spectrum Properties Filter:
-----------------------------
Total Intensity Threshold:  50000
Minimum Peak Count:  15

--------------------------------------------------------------------------------
Processing node 2: SEQUEST
--------------------------------------------------------------------------------

1. Input Data:
-----------------------------
```

Figure 317 gives an example of the output annotated spectrum in the HTML file.

Figure 317. Annotated spectrum linked to the HTML file



The annotation includes the following information:

- The path and name of the raw file
- The numbers of any grouped spectra
- The retention time of the precursor ion, in minutes
- The type of mass spectrometer used to create the data in the raw file
- The fragmentation method used
- The charge state of the precursor ion
- The mass-to-charge ratio (*m/z*) of the precursor ion, in daltons
- The protonated monoisotopic mass (MH+), in daltons
- The match tolerance

Figure 318 gives an example of the output peak list in the HTML file.

Figure 318. Peak list linked to the HTML file

m/z	i	matches
328.88919	0.7	
402.02887	13.5	z (4)(1+)
403.10208	6.8	
473.58258	1.7	
488.40878	2.7	
516.93976	2.9	
518.95996	2.3	
542.70514	1.3	
547.06110	1.6	
586.75372	2.7	
587.43201	2.8	
605.79309	2.5	
606.79248	2.2	
618.44434	0.8	
621.99573	92.8	
622.95319	266.9	[M+3H] (3+)
623.84753	129.5	
624.57965	8.6	
625.29034	3.3	
639.14209	6.0	
640.26099	7.7	
641.07367	11.0	
641.70758	2.1	
653.08392	1.4	
655.37769	3.7	
683.53937	3.1	
700.22180	5.0	
759.26263	10.7	
759.92853	2.6	
775.29376	3.7	
823.68817	4.9	
827.36816	7.8	
828.10382	1.9	
831.81561	2.9	
847.80389	2.4	
855.51782	5.2	
863.45630	4.3	
867.71869	3.4	

Export Annotated Spectra Dialog Box Parameters

Table 96 lists the parameters on the Export Annotated Spectra dialog box.

Table 96. Export Annotated Spectra dialog box parameters (Sheet 1 of 2)

Command or Option	Description
Destination	Specifies the path and name of the PDF file where you want to save the exported annotated spectra.
Criteria	Specifies the type of result data that you want to export: • All PSMs: Exports the annotated spectra of all peptide spectrum matches in the report that are not filtered out. • All PSMs of Checked Proteins: Exports the annotated spectra of the peptide spectrum matches belonging to selected proteins. • With Peptide Grouping Disabled: Makes the following options available only when peptide grouping is not enabled: – Checked PSMs: Exports the annotated spectra of the peptide spectrum matches belonging to selected peptides. • With Peptide Grouping Enabled: Makes the following options available only when peptide grouping is enabled: – Best PSM of All Peptides: Exports the annotated spectra of the best peptide spectrum match of every peptide. – Best PSM of Checked Peptides: Exports the annotated spectra of the best peptide spectrum match for every selected peptide. – All PSMs of Checked Peptides: Exports the annotated spectra of all peptide spectrum matches of every selected peptide.
Criteria, continued	– Best PSM of all Peptides of Checked Proteins: Exports the best peptide spectrum match for every peptide of all selected proteins. The HTML index page contained in the exported zip archive lists the proteins with their annotations and descriptions followed by their peptide spectrum matches.

Table 96. Export Annotated Spectra dialog box parameters (Sheet 2 of 2)

Command or Option	Description
Image Size	Specifies the size of the spectrum images in the PDF file. • Width: Specifies the width of the spectra. The default is 16.0. • Height: Specifies the height of the spectra. The default height is 8.00 centimeters. The Proteome Discoverer application enlarges the spectrum width to the width of the paper. The format of the exported PDF file is A4.
Fragment Match Options	Opens the Options dialog box. For information on setting the fragment match options, see "Setting Global Default Fragment Match Options" on page 53.
Export	Saves the selected annotated spectra in the specified HTML file.
Close	Closes the dialog box without exporting any annotated spectra or applying any of the selected options.

Fragment Match Options Dialog Box

The parameters of the Fragment Match Options dialog box are the same as those of the Options dialog box. For information on these parameters, see "Options Dialog Box Parameters" on page 56.

Exporting Search Results in ProtXML Format

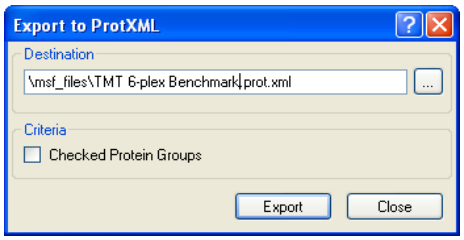
You can export all or filtered proteins in a search report to an XML file in ProtXML format. ProtXML is a data format for storing and exchanging protein identifications. The exported data includes the reporter ion quantitation ratios, precursor ion quantitation ratios, or both, if they are available.

❖ To export search results in ProtXML format

1. Choose **File > Export > To ProtXML**.

   The Export to ProtXML dialog box appears, as shown in Figure 319.

   Figure 319. Export to ProtXML dialog box

2. In the Destination box, type the path and name of the XML file where you will place the exported workflow template, or click the browse button (…) next to the Destination box to browse for the file.
3. (Optional) Select the **Checked Protein Groups** check box to include only selected proteins in the results exported to the XML file.
4. Click **Export**.

   –or–

   To close the dialog box without exporting any results, click **Close**.

Export to ProtXML Dialog Box Parameters

Table 97 lists the parameters on the Export to ProtXML dialog box.

Table 97. Export to ProtXML dialog box parameters

Command or Option	Description
Destination	Specifies the path and name of the XML file to save the proteins in the results to.
Checked Protein Groups	Includes only selected proteins in the results exported to the XML file.
Export	Saves the selected results in the specified XML file.
Close	Closes the Export to ProtXML dialog box without exporting any results.

Exporting Search Results in PepXML Format

You can export peptides that are included in the results of searches performed by all supported search engines in an XML file in PepXML format. This format is an open data format developed by SPC/Institute for Systems Biology for storing, exchanging, and processing peptide sequence assignments from MS/MS scans. It is a common output data format used by many different MS/MS search engines and subsequent peptide analyses.

The Proteome Discoverer application supports the export of search results in PepXML format version 1.14.

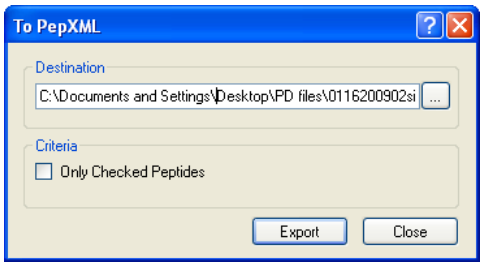
Only peptides and proteins that are not filtered out are exported. You cannot export peptides that are not associated with a protein.

❖ To export search results in PepXML files

1. Open your search results.
2. Choose **File > Export > pepXML**.

   The To PepXML dialog box opens, as shown in Figure 320.

   Figure 320. To PepXML dialog box

3. In the Destination box, type the path and name of the XML file to store the exported peptides in, or click the browse button (...), and in the Save As dialog box, type the name of the XML file to store the peptides in and click **Save**.
4. If you want to export only certain peptides, do the following:

   a. On the report page, place a check mark in the box to the left of the peptide names.

   b. In the To PepXML dialog box, select the **Only Checked Peptides** check box.

   If you do not select the Only Checked Peptides check box, the Proteome Discoverer application generates an XML file that includes all peptides in the search results.
5. Click **Export**.

   –or–

   To close the dialog box without effecting any change, click **Close**.

To PepXML Dialog Box Parameters

Table 98 lists the parameters in the To PepXML dialog box.

Table 98. To PepXML dialog box parameters

Command or Option	Description
Destination	Specifies the path and name of the PepXML file where you want to save the exported peptides.
Only Checked Peptides	Exports all peptides that you selected by putting a check mark next to them. If you do not select the Only Checked Peptides check box, the Proteome Discoverer application exports all peptides included in the search results.
Export	Saves the selected exported peptides in the specified PepXML file.
Close	Closes the dialog box without exporting any peptides.

Copying or Saving a View to an Image

You can copy and save views or graphs to various image formats.

❖ To copy a view or a graph

1. Open a view such as a Chromatogram view.
2. Right-click the view and choose **Copy** from the shortcut menu.

   The view is automatically stored as an image on the Clipboard. You can paste the image into another application.

❖ To save a view in another format

1. Open a view, such as a Chromatogram view.
2. Right-click the view and choose **Save as** from the shortcut menu.
3. Select the image type: **EMF**, **PNG**, **GIF**, **JPG**, **TIFF**, or **BMP**.
4. (Optional) Select the location to store the image in.
5. In the File Name box, type the name of the file.
6. Click **Save**.

The image is saved in the format and location that you selected.

Exporting Protein References to a FASTA File

You can export the protein references and sequences in your search results to a FASTA file.

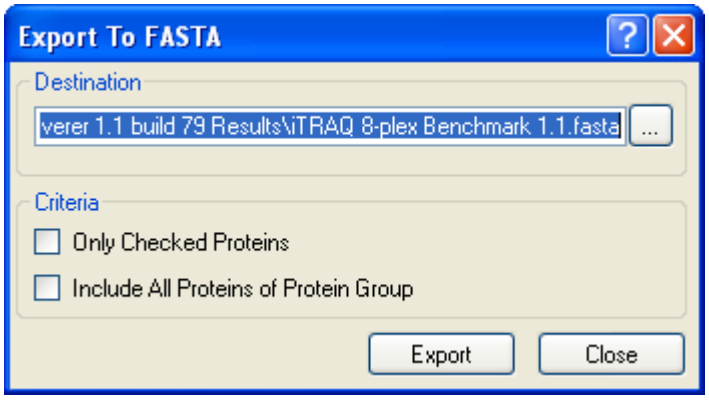
❖ **To export the protein references and sequences in the search results to a FASTA file**

1. On the Proteins or Search Input page, select the proteins that you want to export. Follow the instructions in "Selecting Peptides and Proteins" on page 232 to place a check mark in the box, ☐, next to the proteins that you want to export.

2. Choose **File > Export > To FASTA**.

   The Export to FASTA dialog box appears, as shown in Figure 321.

   Figure 321. Export to FASTA dialog box

3. In the Destination box, type the path and name of the file, or click the browse button (…) to browse to the location where you want to save the FASTA file. Click **Save** in the Save As dialog box.

   The Proteome Discoverer application automatically changes the suffix of the results file from .msf to .fasta.

4. To export only the selected protein references to the FASTA file, select the **Only Checked Proteins** check box.

5. To export all proteins of each protein group to the FASTA file, select the **Include All Proteins of Protein Group** check box.

   If you do not select either the Only Checked Proteins or the Include All Proteins of Protein Group option, the application exports only the master proteins of each protein group. If you do not group proteins, the application exports either all or selected proteins.

6. Click **Export**.

   To close the Export to FASTA dialog box without exporting any results, click **Close**.

Export to FASTA Dialog Box Parameters

Table 99 lists the parameters in the Export to FASTA dialog box.

Table 99. Export to FASTA dialog box parameters

Command or Option	Description
Destination	Specifies the path and name of the FASTA file where you want to save the search results.
Only Checked Proteins	Exports only the master protein of each selected protein group to the FASTA file.
Include All Proteins of Protein Group	Export all proteins of each protein group to the FASTA file.
Export	Saves the selected results in the specified FASTA file.
Close	Closes the Export to FASTA dialog box without exporting any search results.

Exporting Exclusion Mass Lists to Xcalibur

An exclusion mass list is a list of mass-to-charge ratios (*m/z*) that you want to exclude from tandem mass spectrometry (MS/MS).

You can export exclusion mass lists based on your current search results. Exclusion mass lists differ in their usage but have the same basic format.

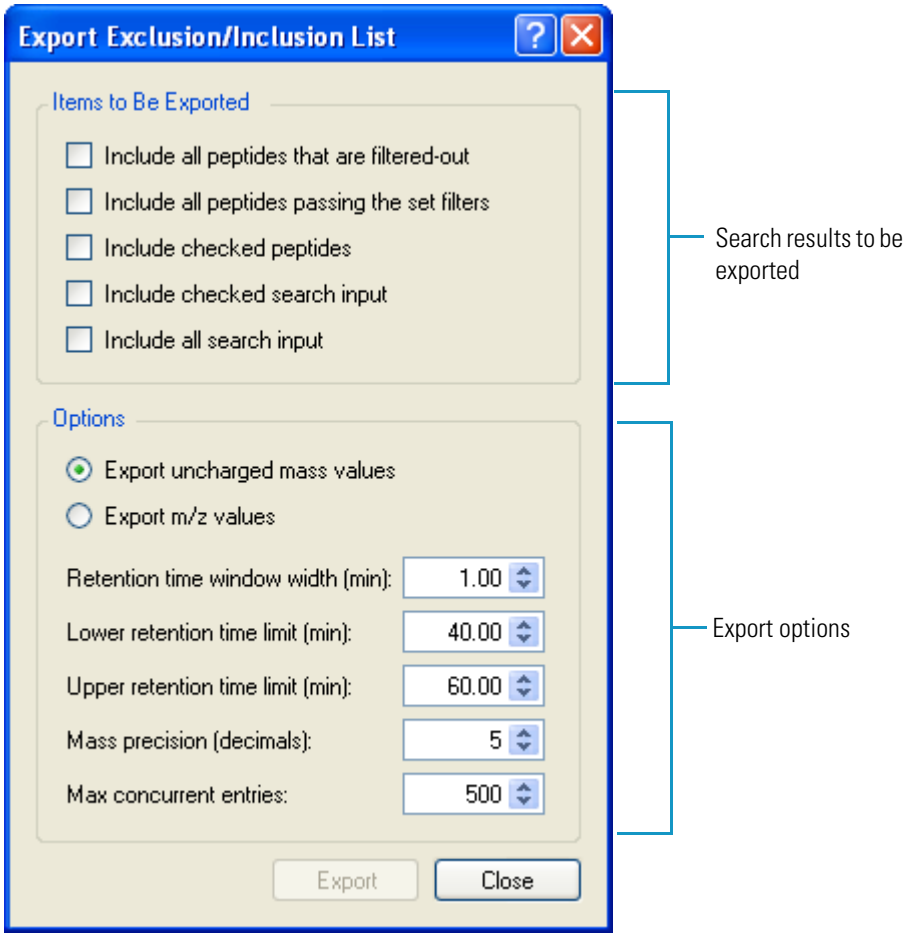
The Export Exclusion/Inclusion List dialog box contains two groups of parameters that you can use to customize the exported mass list. The parameters in the Items to Be Exported area specify which result items should be transferred to the exclusion mass list. The parameters in the Options area customize the format of the resulting exclusion mass list. Use these features to export to a format that can be used in the Xcalibur data system.

❖ To export an exclusion mass list

1. Open your search results.
2. From the Proteome Discoverer application menu bar, choose **File > Export > Xcalibur Exclusion List**.

   The Export Exclusion/Inclusion List dialog box appears, as shown in Figure 322.

Figure 322. Export Exclusion/Inclusion List dialog box



3. In the Items to Be Exported area, specify which result items are to be transferred to the exclusion mass list:

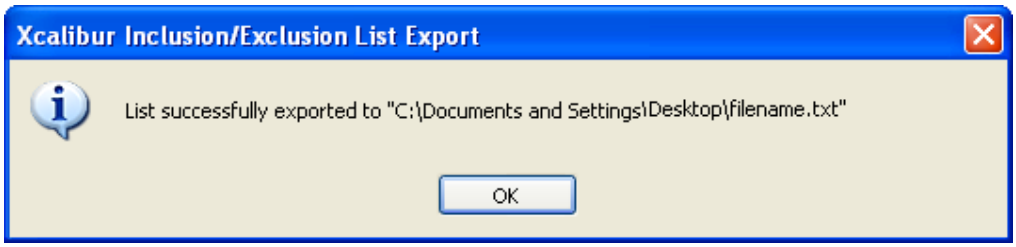
   - Include All Peptides That Are Filtered-Out: Includes the masses of all peptides in the exclusion list that are currently not shown because they do not pass the currently applied result filter set.
   - Include All Peptides Passing the Set Filters: Includes the masses of all currently visible peptides in the exclusion list.
   - Include Checked Peptides: Includes the masses of all peptides in the exclusion list that you selected by putting a check mark on them.
   - Include Checked Search Input: Includes the precursor masses of all spectra in the exclusion list that you selected by putting a check mark next to them on the Search Input page.
   - Include All Search Input: Includes all the precursor masses of all spectra listed on the Search Input page.

4. In the Options area, select either the **Export Uncharged Mass Values** or the **Export m/z Values** option.
   - Export Uncharged Mass Values: Determines whether the masses selected in the Items to Be Exported area are to be exported with their uncharged masses.
   - Export m/z Values: Determines whether the masses selected in the Items to Be Exported area are to be exported with their detected *m/z* values.
5. Select the proper values for your data set for the time value options:
   - Retention Time Window Width (min): Specifies the width of the retention time window, in minutes. Only *n* entries in the exclusion list are allowed per retention time window. *N* is specified by the Max concurrent entries parameter.
   - Lower Retention Time Limit (min): Specifies the minimum retention time for an item to reside on the exclusion list.
   - Upper Retention Time Limit (min): Specifies the maximum retention time for an item to reside on the exclusion list.
   - Mass Precision (decimals): Specifies the number of decimals for each mass that appears in the exclusion list.
   - Max Concurrent Entries: Specifies the maximum number of entries per given retention time window. The width of the window is specified by the Retention time window width parameter.
6. Click **Export**.

   The Save As dialog box opens.
7. Type the file name for the exported list and click **Save**.

   A confirmation message appears, as shown in Figure 323.

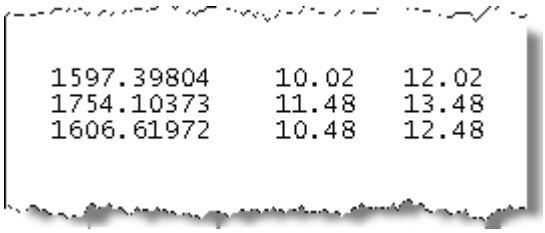
   Figure 323. Export confirmation message



8. Click **OK**.
9. Use a standard text editor to view the resulting exclusion mass list.

   An example of this list is shown in Figure 324.

Figure 324. Exported exclusion mass list example



Export Exclusion/Inclusion List Dialog Box Parameters

Table 100 lists the parameters on the Export Exclusion/Inclusion List dialog box.

Table 100. Export Exclusion/Inclusion List dialog box parameters (Sheet 1 of 2)

Parameter	Description
Include All Peptides That Are Filtered-Out	Includes the masses of all peptides in the exclusion list that are currently not shown because they do not pass the currently applied result filter set.
Include All Peptides Passing the Set Filters	Includes the masses of all currently visible peptides in the exclusion list.
Include Checked Peptides	Includes the masses of all peptides in the exclusion list that you selected by putting a check mark next to them.
Include Checked Search Input	Includes the precursor masses of all spectra in the exclusion list that you selected by putting a check mark next to them on the Search Input page.
Include All Search Input	Includes all the precursor masses of all spectra listed on the Search Input page.
Export Uncharged Mass Values	Determines whether the masses selected in the Items to Be Exported area are to be exported with their uncharged masses.
Export m/z Values	Determines whether the masses selected in the Items to Be Exported area are to be exported with their detected *m/z* values.
Retention Time Window Width (min)	Specifies the width of the retention-time window, in minutes. Only *n* entries in the exclusion list are allowed per retention time window. *N* is specified by the Max concurrent entries parameter.
Lower Retention Time Limit (min)	Specifies the minimum retention time for an item to reside on the exclusion list.
Upper Retention Time Limit (min)	Specifies the maximum retention time value for an item to reside on the exclusion list.

Table 100. Export Exclusion/Inclusion List dialog box parameters (Sheet 2 of 2)

Parameter	Description
Mass Precision (decimals)	Specifies the number of decimals for each mass that appears in the exclusion list.
Max Concurrent Entries	Specifies the maximum number of entries per given retention time window. The width of the window is specified by the retention-time window-width parameter.

Exporting Search Results to Excel

You can copy columns of data from the Proteins, Peptides, and Search Input pages so that you can use them in presentations or Excel spreadsheets.

The Proteome Discoverer application supports Excel 2007, which has a maximum of approximately one million rows. It also supports Excel 97-2003, which supports 65536 rows, but Excel 2007 is the default.

Note When exporting files with large amounts of data, you might receive out-of-memory errors even if you use Excel 2007. In this case, exporting to a tab-delimited text format is a good alternative. See "Exporting Search Results to a Tab-Delimited TXT File" on page 609 for instructions.

- To copy data columns without the headers
- To copy data with the headers
- To copy data to an Excel Workbook

❖ To copy data columns without the headers

1. To select contiguous rows, click the first row that you want to copy, hold down the SHIFT key, and then click the last row that you want to copy.

   To select non-contiguous rows, hold down the CTRL key and click each separate row.
2. Right-click and choose **Copy**.
3. Copy the data into the appropriate document.

❖ To copy data with the headers

1. To select contiguous rows, click the row that you want to copy, hold down the SHIFT key, and then click the last row that you want to copy.

   To select non-contiguous rows, hold down the CTRL key and click each row that you want to copy.
2. Right-click and choose **Copy With Column Headers**.
3. Copy the data into the appropriate document.

❖ To copy data to an Excel Workbook

1. To select contiguous rows, click the first row that you want to copy, hold down the SHIFT key, and then click the last row that you want to copy.

   To select non-contiguous rows, hold down the CTRL key and click each row that you want to copy.

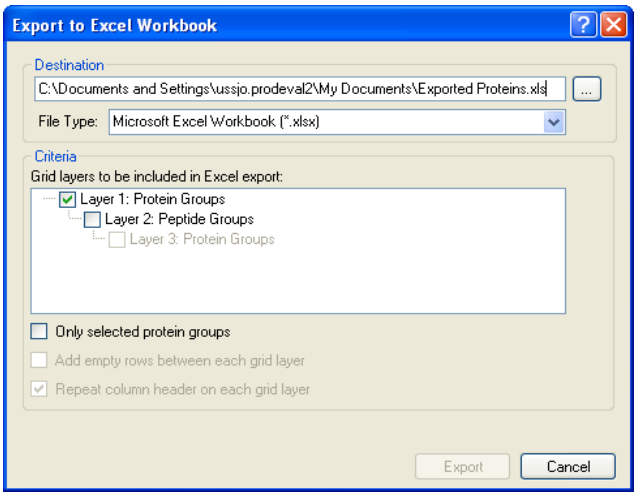
2. Follow the instructions in "Selecting Peptides and Proteins" on page 232 to place a check mark in the box, ☐, next to the peptides or proteins that you want to export.

3. Right-click and choose **Export to Excel Workbook**.

   Note To copy data to an Excel spreadsheet, you must have the Excel software installed.

   The Export to Excel Workbook dialog box now appears, as shown in Figure 325.

   Figure 325. Export to Excel Workbook dialog box



4. In the Destination box, type the path and name of the XLS file, or click the browse button (...) to browse for the file.

5. From the File Type menu, select either **Microsoft Excel Workbook (*.xlsx)** or **Microsoft Excel 97 - 2003 Workbook (*.xls)**.

6. In the Criteria area, select the layers to be included in the Excel spreadsheet.

7. To export only the proteins that you selected in the MSF file, select the **Only Selected Proteins** check box.

8. To add a blank row between each row of data in the Excel spreadsheet for clarity, select the **Add Empty Rows between Each Grid Layer** check box.

9. To add the column header to each level or layer in the Excel spreadsheet, select the **Repeat Column Header on Each Grid Layer** check box.

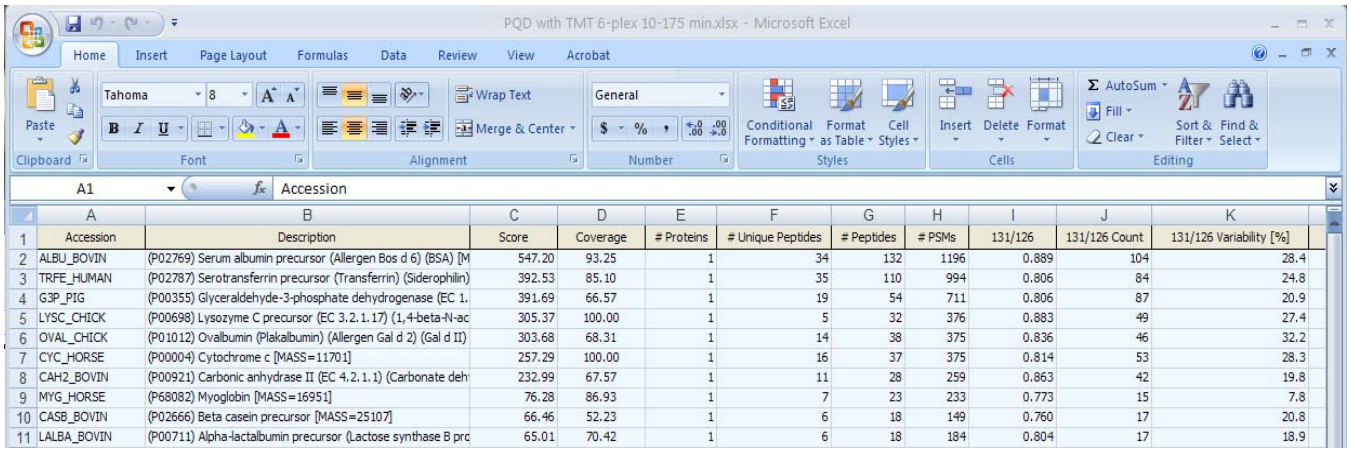
10. Click **Export**.

    A meter above the Export button indicates the progress of the exportation.

11. Go to the directory where the Excel spreadsheet is stored and open it.

If you export just the protein layer, the Excel spreadsheet resembles the example in Figure 326.

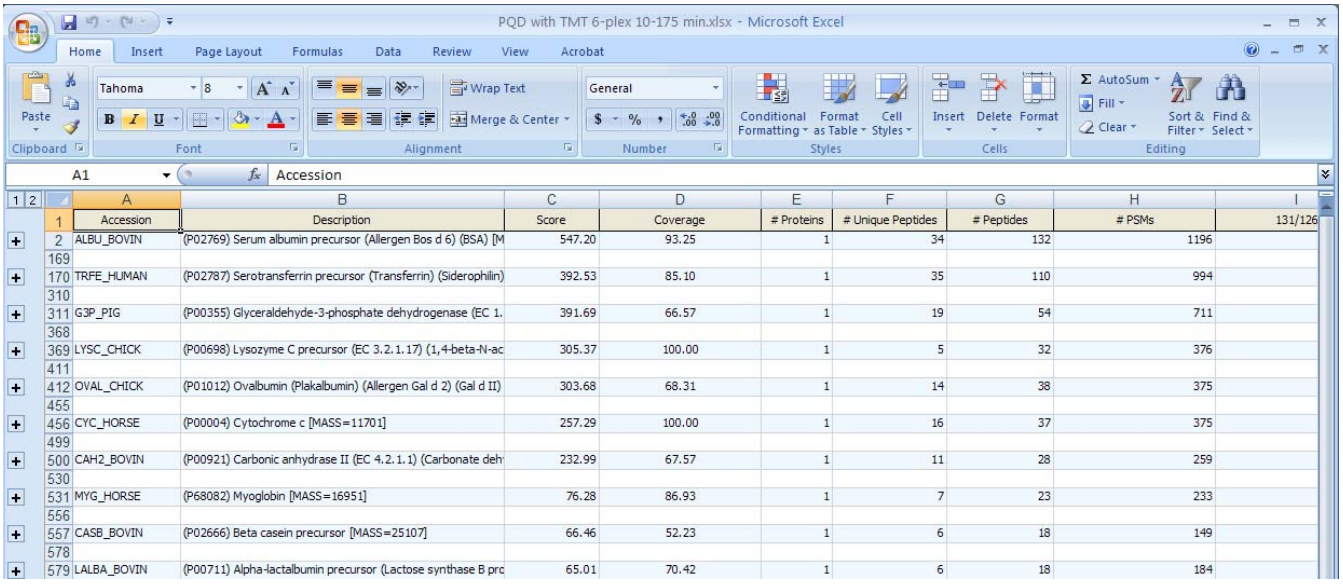
Figure 326. Exporting only the protein layer



	A	B	C	D	E	F	G	H	I	J	K
1	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	131/126	131/126 Count	131/126 Variability [%]
2	ALBU_BOVIN	(P02769) Serum albumin precursor (Allergen Bos d 6) (BSA) [M	547.20	93.25	1	34	132	1196	0.889	104	28.4
3	TRFE_HUMAN	(P02787) Serotransferrin precursor (Transferrin) (Siderophilin)	392.53	85.10	1	35	110	994	0.806	84	24.8
4	G3P_PIG	(P00355) Glyceraldehyde-3-phosphate dehydrogenase (EC 1.	391.69	66.57	1	19	54	711	0.806	87	20.9
5	LYSC_CHICK	(P00698) Lysozyme C precursor (EC 3.2.1.17) (1,4-beta-N-ac	305.37	100.00	1	5	32	376	0.883	49	27.4
6	OVAL_CHICK	(P01012) Ovalbumin (Plakalbumin) (Allergen Gal d 2) (Gal d II)	303.68	68.31	1	14	38	375	0.836	46	32.2
7	CYC_HORSE	(P00004) Cytochrome c [MASS=11701]	257.29	100.00	1	16	37	375	0.814	53	28.3
8	CAH2_BOVIN	(P00921) Carbonic anhydrase II (EC 4.2.1.1) (Carbonate deh	232.99	67.57	1	11	28	259	0.863	42	19.8
9	MYG_HORSE	(P68082) Myoglobin [MASS=16951]	76.28	86.93	1	7	23	233	0.773	15	7.8
10	CASB_BOVIN	(P02666) Beta casein precursor [MASS=25107]	66.46	52.23	1	6	18	149	0.760	17	20.8
11	LALBA_BOVIN	(P00711) Alpha-lactalbumin precursor (Lactose synthase B prc	65.01	70.42	1	6	18	184	0.804	17	18.9

If you chose Add Empty Rows between Each Grid Layer, the Excel spreadsheet resembles the example in Figure 327.

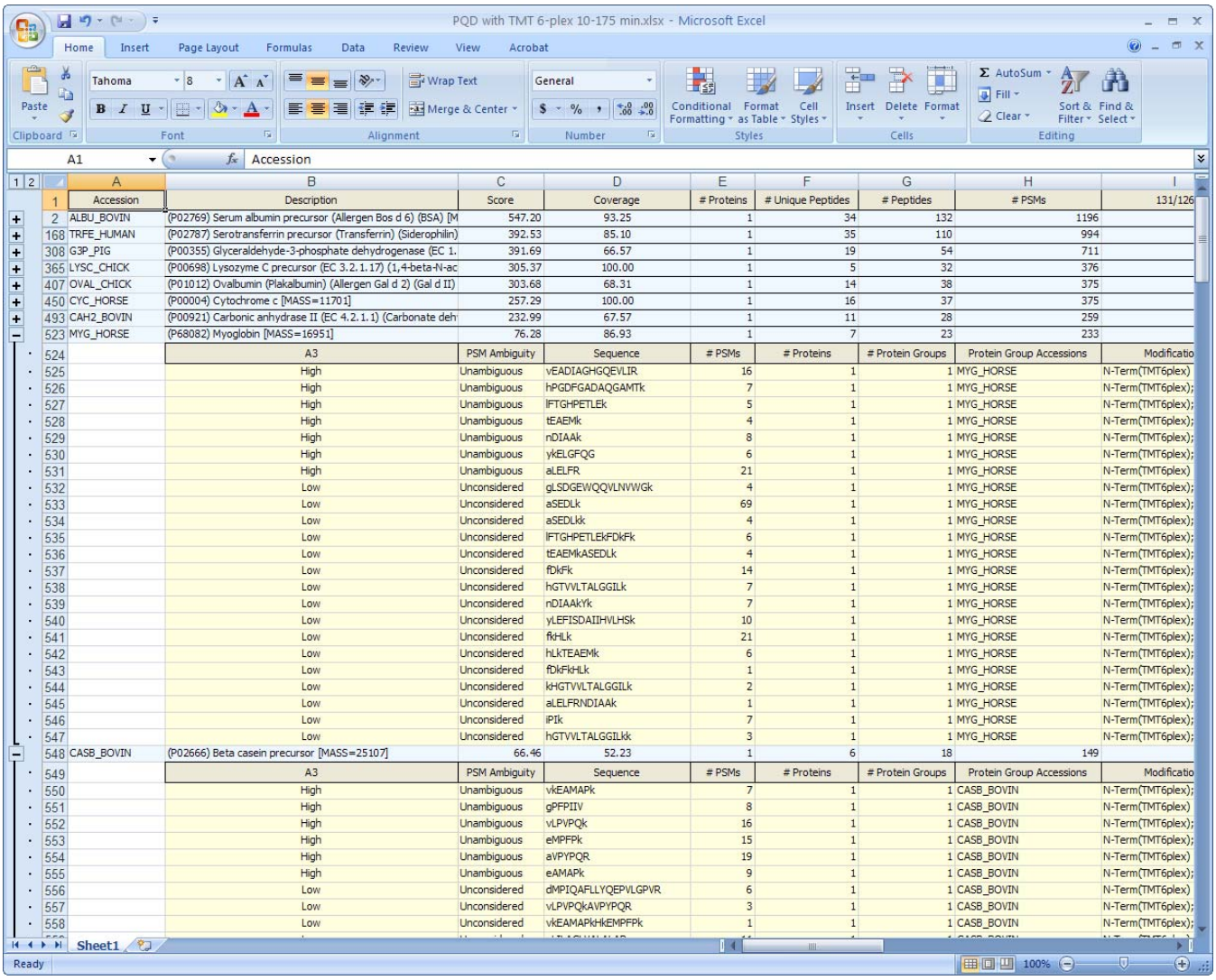
Figure 327. Results of Add Empty Rows between Each Grid Layer option



	A	B	C	D	E	F	G	H	I
1	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	131/126
2	ALBU_BOVIN	(P02769) Serum albumin precursor (Allergen Bos d 6) (BSA) [M	547.20	93.25	1	34	132	1196	
169									
170	TRFE_HUMAN	(P02787) Serotransferrin precursor (Transferrin) (Siderophilin)	392.53	85.10	1	35	110	994	
310									
311	G3P_PIG	(P00355) Glyceraldehyde-3-phosphate dehydrogenase (EC 1.	391.69	66.57	1	19	54	711	
368									
369	LYSC_CHICK	(P00698) Lysozyme C precursor (EC 3.2.1.17) (1,4-beta-N-ac	305.37	100.00	1	5	32	376	
411									
412	OVAL_CHICK	(P01012) Ovalbumin (Plakalbumin) (Allergen Gal d 2) (Gal d II)	303.68	68.31	1	14	38	375	
455									
456	CYC_HORSE	(P00004) Cytochrome c [MASS=11701]	257.29	100.00	1	16	37	375	
499									
500	CAH2_BOVIN	(P00921) Carbonic anhydrase II (EC 4.2.1.1) (Carbonate deh	232.99	67.57	1	11	28	259	
530									
531	MYG_HORSE	(P68082) Myoglobin [MASS=16951]	76.28	86.93	1	7	23	233	
556									
557	CASB_BOVIN	(P02666) Beta casein precursor [MASS=25107]	66.46	52.23	1	6	18	149	
578									
579	LALBA_BOVIN	(P00711) Alpha-lactalbumin precursor (Lactose synthase B prc	65.01	70.42	1	6	18	184	

If you chose Repeat Column Header on Each Grid Layer, the Excel spreadsheet resembles the example in Figure 328.

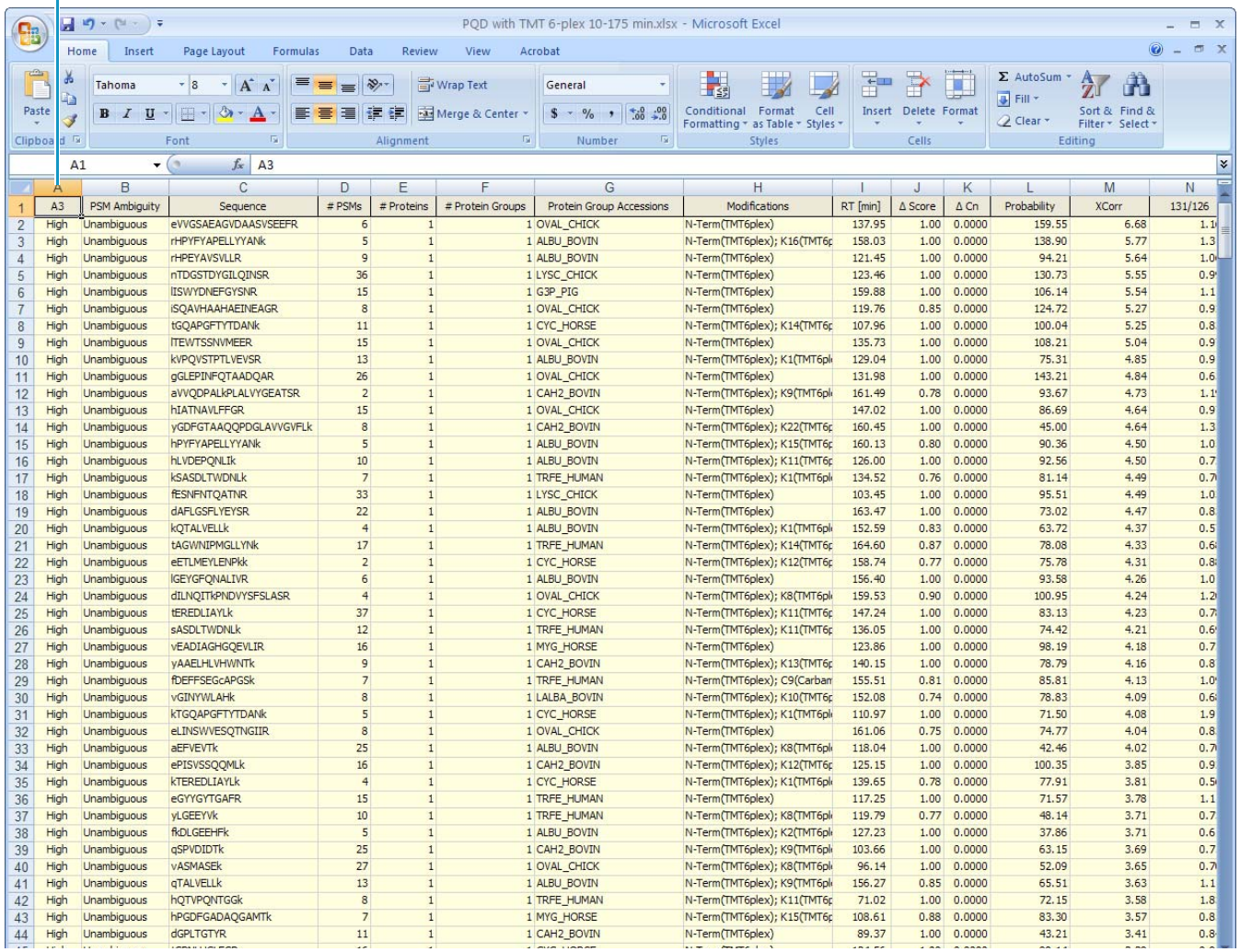
Figure 328. Results of Repeat Column Header on Each Grid Layer option



When you export data from the Peptide results pages, the Proteome Discoverer application automatically exports peptide confidences as well. They appear in the Confidence column in the Excel spreadsheet, shown in Figure 329. This column includes High, Medium, and Low labels that correspond to the graphical confidence indicators (the green, yellow, and red circles attached to each peptide match) on the Proteome Discoverer result pages.

Figure 329. Excel spreadsheet with Confidence column

Confidence column



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	A3	PSM Ambiguity	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	RT [min]	Δ Score	Δ Cn	Probability	XCorr	131/126
2	High	Unambiguous	eVVGSAEAGVDAASVSEEFR	6	1	1	OVAL_CHICK	N-Term(TMT6plex)	137.95	1.00	0.0000	159.55	6.68	1.1
3	High	Unambiguous	rHPYFYAPELLYYANk	5	1	1	ALBU_BOVIN	N-Term(TMT6plex); K16(TMT6p	158.03	1.00	0.0000	138.90	5.77	1.3
4	High	Unambiguous	rHPEYAVSVLLR	9	1	1	ALBU_BOVIN	N-Term(TMT6plex)	121.45	1.00	0.0000	94.21	5.64	1.0
5	High	Unambiguous	nTDGSTDYGILQINSR	36	1	1	LYSC_CHICK	N-Term(TMT6plex)	123.46	1.00	0.0000	130.73	5.55	0.9
6	High	Unambiguous	IISWYDNEFGYSNR	15	1	1	G3P_PIG	N-Term(TMT6plex)	159.88	1.00	0.0000	106.14	5.54	1.1
7	High	Unambiguous	iSQAVHAAHAEINEAGR	8	1	1	OVAL_CHICK	N-Term(TMT6plex)	119.76	0.85	0.0000	124.72	5.27	0.9
8	High	Unambiguous	tGQAPGFTYTDANk	11	1	1	CYC_HORSE	N-Term(TMT6plex); K14(TMT6p	107.96	1.00	0.0000	100.04	5.25	0.8
9	High	Unambiguous	lTEWTSSNVMEER	15	1	1	OVAL_CHICK	N-Term(TMT6plex)	135.73	1.00	0.0000	108.21	5.04	0.9
10	High	Unambiguous	kVPQVSTPTLVEVSR	13	1	1	ALBU_BOVIN	N-Term(TMT6plex); K1(TMT6pl	129.04	1.00	0.0000	75.31	4.85	0.9
11	High	Unambiguous	gGLEPINFQTAADQAR	26	1	1	OVAL_CHICK	N-Term(TMT6plex)	131.98	1.00	0.0000	143.21	4.84	0.6
12	High	Unambiguous	aVVQDPALkPLALVYGEATSR	2	1	1	CAH2_BOVIN	N-Term(TMT6plex); K9(TMT6pl	161.49	0.78	0.0000	93.67	4.73	1.1
13	High	Unambiguous	hIATNAVLFFGR	15	1	1	OVAL_CHICK	N-Term(TMT6plex)	147.02	1.00	0.0000	86.69	4.64	0.9
14	High	Unambiguous	yGDFGTAAQQPDGLAVVGVFLk	8	1	1	CAH2_BOVIN	N-Term(TMT6plex); K22(TMT6p	160.45	1.00	0.0000	45.00	4.64	1.3
15	High	Unambiguous	hPYFYAPELLYYANk	5	1	1	ALBU_BOVIN	N-Term(TMT6plex); K15(TMT6p	160.13	0.80	0.0000	90.36	4.50	1.0
16	High	Unambiguous	hLVDEPQNLIk	10	1	1	ALBU_BOVIN	N-Term(TMT6plex); K11(TMT6p	126.00	1.00	0.0000	92.56	4.50	0.7
17	High	Unambiguous	kSASDLTWDNLk	7	1	1	TRFE_HUMAN	N-Term(TMT6plex); K1(TMT6pl	134.52	0.76	0.0000	81.14	4.49	0.7
18	High	Unambiguous	fESNFNTQATNR	33	1	1	LYSC_CHICK	N-Term(TMT6plex)	103.45	1.00	0.0000	95.51	4.49	1.0
19	High	Unambiguous	dAFLGSFLYEYSR	22	1	1	ALBU_BOVIN	N-Term(TMT6plex)	163.47	1.00	0.0000	73.02	4.47	0.8
20	High	Unambiguous	kQTALVELLk	4	1	1	ALBU_BOVIN	N-Term(TMT6plex); K1(TMT6pl	152.59	0.83	0.0000	63.72	4.37	0.5
21	High	Unambiguous	tAGWNIPMGLLYNk	17	1	1	TRFE_HUMAN	N-Term(TMT6plex); K14(TMT6p	164.60	0.87	0.0000	78.08	4.33	0.6
22	High	Unambiguous	eETLMEYLENPkk	2	1	1	CYC_HORSE	N-Term(TMT6plex); K12(TMT6p	158.74	0.77	0.0000	75.78	4.31	0.8
23	High	Unambiguous	lGEYGFQNALIVR	6	1	1	ALBU_BOVIN	N-Term(TMT6plex)	156.40	1.00	0.0000	93.58	4.26	1.0
24	High	Unambiguous	dILNQITkPNDVYSFSLASR	4	1	1	OVAL_CHICK	N-Term(TMT6plex); K8(TMT6pl	159.53	0.90	0.0000	100.95	4.24	1.2
25	High	Unambiguous	tEREDLIAYLk	37	1	1	CYC_HORSE	N-Term(TMT6plex); K11(TMT6p	147.24	1.00	0.0000	83.13	4.23	0.7
26	High	Unambiguous	sASDLTWDNLk	12	1	1	TRFE_HUMAN	N-Term(TMT6plex); K11(TMT6p	136.05	1.00	0.0000	74.42	4.21	0.6
27	High	Unambiguous	vEADIAGHGQEVLIR	16	1	1	MYG_HORSE	N-Term(TMT6plex)	123.86	1.00	0.0000	98.19	4.18	0.7
28	High	Unambiguous	yAAELHLVHWNTk	9	1	1	CAH2_BOVIN	N-Term(TMT6plex); K13(TMT6p	140.15	1.00	0.0000	78.79	4.16	0.8
29	High	Unambiguous	fDEFFSEGcAPGSk	7	1	1	TRFE_HUMAN	N-Term(TMT6plex); C9(Carbam	155.51	0.81	0.0000	85.81	4.13	1.0
30	High	Unambiguous	vGINYWLAHk	8	1	1	LALBA_BOVIN	N-Term(TMT6plex); K10(TMT6p	152.08	0.74	0.0000	78.83	4.09	0.6
31	High	Unambiguous	kTGQAPGFTYTDANk	5	1	1	CYC_HORSE	N-Term(TMT6plex); K1(TMT6pl	110.97	1.00	0.0000	71.50	4.08	1.9
32	High	Unambiguous	eLINSWVESQTNGIIR	8	1	1	OVAL_CHICK	N-Term(TMT6plex)	161.06	0.75	0.0000	74.77	4.04	0.8
33	High	Unambiguous	aEFVEVTk	25	1	1	ALBU_BOVIN	N-Term(TMT6plex); K8(TMT6pl	118.04	1.00	0.0000	42.46	4.02	0.7
34	High	Unambiguous	ePISVSSQQMLk	16	1	1	CAH2_BOVIN	N-Term(TMT6plex); K12(TMT6p	125.15	1.00	0.0000	100.35	3.85	0.9
35	High	Unambiguous	kTEREDLIAYLk	4	1	1	CYC_HORSE	N-Term(TMT6plex); K1(TMT6pl	139.65	0.78	0.0000	77.91	3.81	0.5
36	High	Unambiguous	eGYYGYTGAFR	15	1	1	TRFE_HUMAN	N-Term(TMT6plex)	117.25	1.00	0.0000	71.57	3.78	1.1
37	High	Unambiguous	yLGEEYVk	10	1	1	TRFE_HUMAN	N-Term(TMT6plex); K8(TMT6pl	119.79	0.77	0.0000	48.14	3.71	0.7
38	High	Unambiguous	fkDLGEEHFk	5	1	1	ALBU_BOVIN	N-Term(TMT6plex); K2(TMT6pl	127.23	1.00	0.0000	37.86	3.71	0.6
39	High	Unambiguous	qSPVDIDTk	25	1	1	CAH2_BOVIN	N-Term(TMT6plex); K9(TMT6pl	103.66	1.00	0.0000	63.15	3.69	0.7
40	High	Unambiguous	vASMASEk	27	1	1	OVAL_CHICK	N-Term(TMT6plex); K8(TMT6pl	96.14	1.00	0.0000	52.09	3.65	0.7
41	High	Unambiguous	qTALVELLk	13	1	1	ALBU_BOVIN	N-Term(TMT6plex); K9(TMT6pl	156.27	0.85	0.0000	65.51	3.63	1.1
42	High	Unambiguous	hQTVPQNTGGk	8	1	1	TRFE_HUMAN	N-Term(TMT6plex); K11(TMT6p	71.02	1.00	0.0000	72.15	3.58	1.8
43	High	Unambiguous	hPGDFGADAQGAMTk	7	1	1	MYG_HORSE	N-Term(TMT6plex); K15(TMT6p	108.61	0.88	0.0000	83.30	3.57	0.8
44	High	Unambiguous	dGPLTGTYR	11	1	1	CAH2_BOVIN	N-Term(TMT6plex)	89.37	1.00	0.0000	43.21	3.41	0.8

Export to Excel Workbook Dialog Box Parameters

Table 101 lists the parameters in the Export to Excel Workbook dialog box.

Table 101. Export to Excel Workbook dialog box parameters

Command or Option	Description
Destination	Specifies the path and name of the XLS (Excel spreadsheet) file to export the search results to.
File Type	Specifies the type of file to export the search results to: • (Default) Microsoft Excel Workbook (*xlsx): Exports search results to Excel 2007, which has a maximum of approximately one million rows. • Microsoft 97 - 2003 Workbook (*xls): Exports search results to Excel 97 - 2003, which has a maximum of 65536 rows.
Criteria	Specifies the layers to be included in the Excel spreadsheet.
Only Selected Proteins	Exports only the proteins that you selected in the MSF file.
Add Empty Rows between Each Grid Layer	Adds a blank row between each row of data in the Excel spreadsheet.
Repeat Column Header on Each Grid Layer	Adds the column header to each level or layer in the Excel spreadsheet.
Enter	Exports the selected data in the Proteins, Peptides, and Search Input pages of the MSF file to an XLS (Excel spreadsheet) file.

Exporting Search Results to a Tab-Delimited TXT File

You can export search results to a tab-delimited TXT file. The TXT file format is a simple text format that stores tabular data and is widely used to exchange data between different computer programs. Lines in a file represent table rows, and special characters (tabs) in a line separate the fields in a row. You can export the proteins, peptides, and search inputs in an open MSF report to a file in TXT format. The Proteome Discoverer application exports each page (Proteins, Peptides, or Search Input) of the MSF file in a separate file. It exports all data columns in the MSF file, whether a particular column is displayed or not. The application uses tabs to separate the fields in a row in the exported files.

The name of the exported file combines the destination file name, the name of the page (_proteingroups, _proteins, _peptidegroups, _psms, or _searchinput), and a .txt extension. Here are some examples:

E:\TestData\MultiConsensus_proteins.txt

E:\TestData\MultiConsensus_peptides.txt

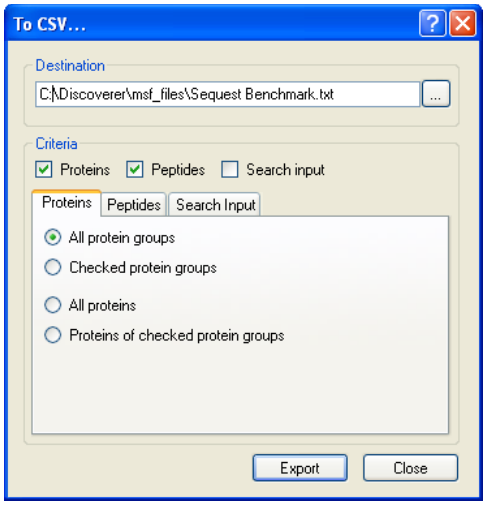
E:\TestData\MultiConsensus_searchinput.txt

❖ To export search results to a tab-delimited TXT file

1. Open your search results in the MSF file.
2. Choose **File > Export > To CSV**.

   The To CSV dialog box opens, as shown in Figure 330.

   Figure 330. Proteins page of the To CSV dialog box



3. In the Destination box, specify the path and name of the file where you want to save the search results.
4. In the Criteria area, select the check box or boxes corresponding to the type of result data that you want to export. You can choose Proteins, Peptides, Search Input, or all three:
   - (Default) Proteins: Exports the protein data listed on the Proteins page of the MSF file. For more information on this page, see "Using the Proteins Page" on page 237.
   - (Default) Peptides: Exports the peptide data listed on the Peptides page of the MSF file. For more information on this page, see"Using the Peptides Page" on page 288.
   - Search Input: Exports the precursor ion data listed on the Search Input page of the MSF file. For more information on this page, see "Using the Search Input Page" on page 346.

Note The Proteome Discoverer application automatically appends the _proteingroups, _proteins, _peptidegroups, _psms, and _searchinput suffixes to the .txt files for the five different data types exported.

5. Select the export options for each type of data that you want to export:

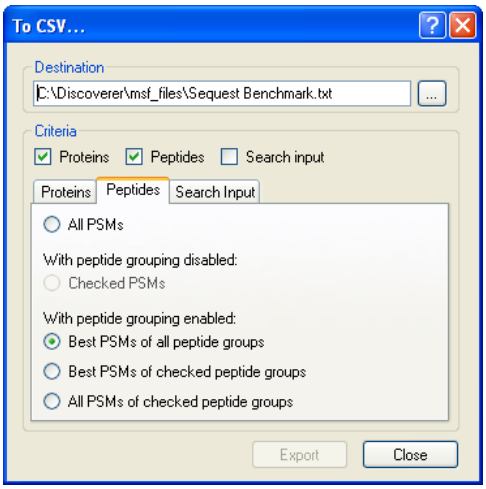
   - If you selected Proteins in the Criteria area, click the **Proteins** tab in the dialog box and select your options.

     The Proteins tab is shown in Figure 330. For information about the options on this page, see Table 102.

   - If you selected Peptides in the Criteria area, click the **Peptides** tab in the dialog box and select your options.

     The Peptides page is shown in Figure 331. For information about the options on this page, see Table 102.

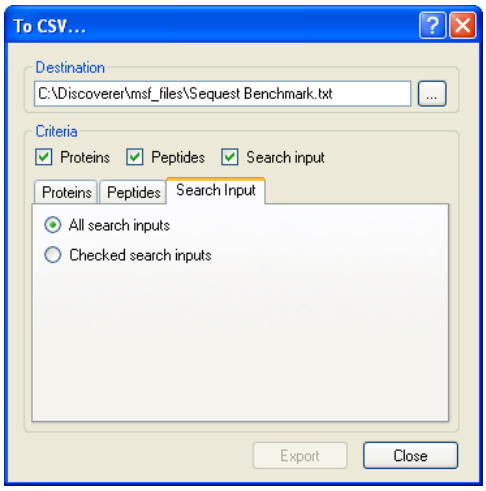
     Figure 331. Peptides page of the To CSV dialog box

   - If you selected Search Input in the Criteria area, click the **Search Input** tab in the dialog box and select your options.

     The Search Input page is shown in Figure 332. For information about the options on this page, see Table 102.

Figure 332. Search Input page of the To CSV dialog box



6. Click **Export**.

The Proteome Discoverer application saves the selected data in a TXT file. You can drag this file to an open Microsoft Excel application to open it.

To CSV Dialog Box Parameters

Table 102 lists the parameters in the To CSV dialog box.

Table 102. To CSV dialog box parameters (Sheet 1 of 2)

Command or Option	Description
Destination	Specifies the path and name of the tab-delimited TXT (comma-separated values) file where you want to save the search results.
Criteria	Specifies the type of result data that you want to export: • (Default) Proteins: Exports the protein data listed on the Proteins page of the MSF file. For more information on this page, see "Using the Proteins Page" on page 237. • (Default) Peptides: Exports the peptide data listed on the Peptides page of the MSF file. For more information on this page, see"Using the Peptides Page" on page 288. • Search Input: Exports the precursor ion data listed on the Search Input page of the MSF file. For more information on this page, see "Using the Search Input Page" on page 346.

Table 102. To CSV dialog box parameters (Sheet 2 of 2)

Command or Option	Description
Proteins page	Specifies the type of protein data to export: • (Default) All Protein Groups: Exports the data of all protein groups. • Checked Protein Groups: Exports the data of selected protein groups. • All Proteins: Exports the data of all proteins. • Proteins of Checked Protein Groups: Exports the data of the proteins of selected protein groups.
Peptides page	Specifies the type of peptide data to export: • All PSMs: Exports the data of all peptide spectrum matches that are not filtered out. • With Peptide Grouping Enabled: Makes the following options available only when peptide grouping is enabled: – (Default) Best PSM of All Peptides: Exports the data of the best peptide spectrum match of every peptide. – Best PSM of Checked Peptides: Exports the data of the best peptide spectrum match for every selected peptide. – All PSMs of Checked Peptides: Exports the data of all peptide spectrum matches of every selected peptide.
Search Input page	Specifies the type of search input data to export: • (Default) All Search Inputs: Exports the data of all search inputs. • Checked Search Inputs: Exports the data of selected search inputs.
Export	Saves the selected data in the specified TXT file.
Close	Closes the To CSV dialog box without exporting any data.

Reference

This chapter describes the menu commands on the Proteome Discoverer application. The menus are listed in the order in which they appear on the toolbar.

Contents

- File Menu
- Search Report Menu
- Quantification Menu
- Processing Menu
- Workflow Editor Menu
- Administration Menu
- Tools Menu
- Window Menu
- Help Menu
- Shortcut Menus
- Workflow Editor Nodes

File Menu

The File menu contains the following commands.

Table 103. File menu (Sheet 1 of 3)

Command	Function	Shortcut	Toolbar icon
File > Open Report	Opens a report.	CTRL+O	
File > Save Report	Saves a modified MSF file, specifically: • The peptide confidence and FDR settings of the Peptide Confidence page • The information in the Annotation column of the Peptides page • The information resulting from the setting that you selected by right-clicking on the peptide rows and choosing the Include/Exclude Peptide(s) from Protein Quantification command (see Table 112 for information on the settings available with this command) This command is only available if you have applied a filter to the MSF file or changed the quantification method parameters.	CTRL+S	
File > Close	Closes a report without saving any modifications, such as applied filters. The Proteome Discoverer application remains open after it closes the report. This command is only available if an MSF file is open.		
File > Import Search Results from SRF	Activates the SRF File Import Wizard so that you can import search results from BioWorks 3.3.1 that are in SRF format. For instructions on importing an SRF file, see "Migrating from BioWorks to the Proteome Discoverer Application" on page 715.		

Table 103. File menu (Sheet 2 of 3)

Command	Function	Shortcut	Toolbar icon
File > Export	Opens a menu containing commands that export spectra; export annotated spectra; export search result rows to a ProtXML file, a PepXML file, a tab-delimited TXT file; export the proteins in the search results to a FASTA file; and export an Xcalibur exclusion mass list to an Excel file.		
File > Export > Spectra	Opens the Export Spectra dialog box, shown in Figure 307 on page 578, so that you can export specified spectra. For information about exporting spectra, see "Exporting Spectra" on page 578.		
File > Export > Annotated Spectra	Opens the Export Annotated Spectra dialog box, shown in Figure 309 on page 583, so that you can export spectra annotated with observed masses and fragment assignments in a ZIP file. For information about exporting annotated spectra, see "Exporting Annotated Spectra for Selected Proteins or Peptide Groups" on page 582.		
File > Export > To CSV	Opens the To CSV dialog box, shown in Figure 330 on page 610 so that you can export search results to a tab-delimited TXT file. For information about exporting data to a tab-delimited TXT file, see "Exporting Search Results to a Tab-Delimited TXT File" on page 609.		
File > Export > To PepXML	Opens the To PepXML dialog box, shown in Figure 320 on page 597, so that you can export peptides that are included in the results of searches performed by the SEQUEST and Mascot search engines.		

Table 103. File menu (Sheet 3 of 3)

Command	Function	Shortcut	Toolbar icon
File > Export > To ProtXML	Opens the Export to ProtXML dialog box, shown in Figure 319 on page 596, so that you can export selected search result rows into an XML file. For information about exporting results to a ProtXML file, see "Exporting Search Results in ProtXML Format" on page 596.		
File > Export > To FASTA	Opens the Export to FASTA dialog box, shown in Figure 321 on page 599, so that you can export the protein references in the search results to a FASTA file. For information about exporting proteins to a FASTA file, see "Exporting Protein References to a FASTA File" on page 599.		
File > Export > Xcalibur Exclusion List	Opens the Export Exclusion/Inclusion List dialog box, shown in Figure 322 on page 601, so that you can export an exclusion mass list, which is a list of mass-to-charge ratios (*m/z*) that you want to exclude from tandem mass spectrometry (MS/MS). For more information about exporting the exclusion and inclusion mass lists, see "Exporting Exclusion Mass Lists to Xcalibur" on page 600.		
File > Exit	Closes all opened views and reports without saving any unspecified modifications before exiting the Proteome Discoverer application.		

Search Report Menu

The Search Report menu contains the following commands.

Table 104. Search Report menu (Sheet 1 of 4)

Command	Function	Shortcut	Toolbar icon
Search Report > Show Details	Makes the Protein ID Details command available when you select a protein row or makes the Peptide ID Details command available when you select a peptide row.	CTRL+D	
Search Report > Show Protein ID Details	Opens the Protein Identification Details view, shown in Figure 147 on page 246. For information about this view, see "Interpreting Your Results with the Protein Identification Details View" on page 245.	CTRL+D	
Search Report > Show Peptide ID Details	Opens the Peptide Identification Details view, shown in Figure 161 on page 301. For information about the Peptide Identification Details view, see "Interpreting Your Results with the Peptide Identification Details View" on page 299.	CTRL+D	
Search Report > Show Distribution Chart	Opens the Report Item Distribution chart, which displays different aspects of the search results plotted in relationship to each other in scatter plots (Figure 152 on page 259), Venn diagrams (shown in Figure 153 on page 273), and pie charts (as shown in Figure 154 on page 279). For more information about the Report Item Distribution view, see "Interpreting Your Results with the Report Item Distribution Chart" on page 257.	CTRL+SHIFT+D	
Search Report > Show Peptide Consensus View	Opens the Peptide Consensus view, shown in Figure 168 on page 317, which displays the ion and peptide search results data. For more information about the Peptide Consensus view, see "Interpreting Your Results with the Peptide Consensus View" on page 316.		

Table 104. Search Report menu (Sheet 2 of 4)

Command	Function	Shortcut	Toolbar icon
Search Report > Show Protein Group Members	Opens the Protein Group Members view, shown in Figure 156 on page 286, which displays the proteins that compose the selected protein group. For more information about the Protein Group Members view, see "Interpreting Your Results with the Protein Group Members View" on page 285.	CTRL+SHIFT+H	
Search Report > Show Peptides Group Members	Opens the Peptides Group Members view, shown in Figure 186 on page 344, which displays which peptides are included in a peptide group when you enable peptide grouping. For more information about the Peptides Group Members view, see "Interpreting Your Results with the Peptide Group Members View" on page 343.		
Search Report > Show Chromatogram View	Opens the Chromatogram view, shown in Figure 179 on page 331. For more information about the Chromatogram view, see "Interpreting Your Results with the Chromatogram View" on page 330.	CTRL+SHIFT+C	
Search Report > Show Spectrum	Opens the Spectrum view, shown in Figure 181 on page 334, which displays the MS/MS spectrum that was used for the peptide search. For more information about the Spectrum view, see "Interpreting Your Results with the Spectrum View" on page 333.	CTRL+SHIFT+S	
Search Report > Show Fragment Match Spectrum	Opens the Fragment Match Spectrum view, shown in Figure 182 on page 336, which shows the MS/MS spectrum used for the search, annotated with the fragments that have been assigned within a predefined mass tolerance. See "Calculating the Delta Score" on page 345 for information about the Fragment Match Spectrum view.	CTRL+SHIFT+F	

Table 104. Search Report menu (Sheet 3 of 4)

Command	Function	Shortcut	Toolbar icon
Search Report > Show Precursor Isotope Pattern	Opens the Precursor Isotope Pattern view, shown in Figure 183 on page 338, which shows a detailed perspective of the MS scan of the precursor isotope pattern of the selected precursor. For information about the Precursor Isotope Pattern view, see "Interpreting Your Results with the Precursor Isotope Pattern View" on page 337.	CTRL+SHIFT+P	
Search Report > Show Extracted Ion Chromatogram	Opens the Extracted Ion Chromatogram (XIC) view, shown in Figure 185 on page 342, of the precursor mass for the selected peptide. For information about the Extracted Ion Chromatogram view, see "Interpreting Your Results with the Extracted Ion Chromatogram View" on page 340.	CTRL+SHIFT+T	
Search Report > Show Sequence Comparison	Opens the Sequence Comparison view, which displays the complete sequences for different proteins. You can use this view to compare similar proteins that share varying degrees of identical peptide sequences or to compare the complete sequences of different proteins associated with a single peptide. Figure 155 on page 284 gives an example of the Sequence Comparison view. For more information about the Sequence Comparison window, see "Interpreting Your Results with the Sequence Comparison View" on page 283.		
Search Report > Layout	Provides a menu of options to apply and save layouts. For information about these options, see "Creating and Applying Custom Layouts" on page 233.		
Search Report > Apply Layout	Applies a selected layout to a results report.		
Search Report > Apply Default Layout	Applies the default layout to a results report.		
Search Report > Save Layout	Saves the current layout.		

Table 104. Search Report menu (Sheet 4 of 4)

Command	Function	Shortcut	Toolbar icon
Search Report > Save Layout As	Saves the current layout with a new layout name.		
Search Report > Save Layout As Default	Saves the current layout as the default layout.		
Search Report > Restore Factory Default	Returns the default layout to the layout in effect when you first installed the Proteome Discoverer application.		

Quantification Menu

The Quantification menu contains the following commands.

Table 105. Quantification menu (Sheet 1 of 3)

Command	Function	Shortcut	Toolbar icon
Quantification > Edit Quantification Method	Opens the Quantification Method Editor dialog box, shown in Figure 257 on page 493 through Figure 262 on page 501, so that you can change the quantification method of the current report. For more information about changing quantification methods, see "Changing a Quantification Method" on page 520.		
Quantification > Show Peptide Ratios	Opens the Peptide Ratio Distributions chart, shown in Figure 286 on page 551, which displays the set of calculated relative ratios for the peptides associated with the selected protein. For information about calculating peptide ratios, see "Calculating Peptide Ratios" on page 550.	CTRL+SHIFT+R	

Table 105. Quantification menu (Sheet 2 of 3)

Command	Function	Shortcut	Toolbar icon
Quantification > Show Quan Channel Values	Displays a Quantification Channel Values chart: • For reporter ion quantification, displays the absolute intensity of the reporter ions detected for the available quantification channels, as shown in Figure 274 on page 530. • For precursor ion quantification, the chart displays the area of the isotopes detected for the available quantification channels, as shown in Figure 275 on page 531. For more information about displaying quantification channel values, see "Displaying Quantification Channel Values for Reporter Ion Quantification" on page 529 and "Displaying Quantification Channel Values for Precursor Ion Quantification" on page 530.	CTRL+SHIFT+N	

Table 105. Quantification menu (Sheet 3 of 3)

Command	Function	Shortcut	Toolbar icon
Quantification > Show Quantification Spectrum	Displays the Quantification Spectrum chart: • For reporter ion quantification, the chart shows the absolute and corrected intensities of the reporter ions, as shown in Figure 276 on page 534. • For precursor ion quantification, the chart shows the different abundances of the identified Light, Medium, and Heavy isotopic peak patterns used to quantify a peptide, as shown in Figure 280 on page 539. For instructions on displaying the quantification spectrum chart, see "Displaying the Quantification Spectrum Chart for Reporter Ion Quantification" on page 533 and "Displaying the Quantification Spectrum Chart for Precursor Ion Quantification" on page 538.	CTRL+SHIFT+Q	

Processing Menu

The Processing menu contains the following commands.

Table 106. Processing menu (Sheet 1 of 2)

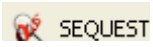
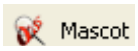
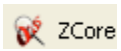
Command	Function	Shortcut	Toolbar icon
Processing > Start SEQUEST Search Wizard	Starts the SEQUEST search wizard. For information about searching with the SEQUEST wizard, see "Starting a New Search by Using the Search Wizards" on page 63.	CTRL+Q	

Table 106. Processing menu (Sheet 2 of 2)

Command	Function	Shortcut	Toolbar icon
Processing > Start Mascot Search Wizard	Starts the Mascot search wizard. For information about searching with the Mascot wizard, see "Starting a New Search by Using the Search Wizards" on page 63.	CTRL+M	
Processing > Start ZCore Search Wizard	Starts the ZCore search wizard. For information about searching with the ZCore wizard, see "Starting a New Search by Using the Search Wizards" on page 63.	CTRL+Z	

Workflow Editor Menu

The Workflow Editor menu contains the following commands.

Table 107. Workflow Editor menu (Sheet 1 of 2)

Command	Function	Shortcut	Toolbar icon
Workflow Editor > New Workflow	Opens a new window in the Workflow Editor, shown in Figure 49 on page 92. For information about creating a new workflow, see "Starting a New Search by Using the Workflow Editor" on page 89. For a description of all the nodes that you can use in a workflow, see "Workflow Editor Nodes" on page 637.		
Workflow Editor > Open From Template	Opens an existing workflow. For information about using this command, see "Opening an Existing Workflow from a Template" on page 109.		
Workflow Editor > Save As Template	Saves a workflow. For information about saving a workflow as a template, see "Creating a Search Workflow" on page 91.		
Workflow Editor > Auto Layout	Automatically adjusts and aligns the connecting arrows and nodes in the workspace in the Workflow Editor. For information about using this command, see "Creating a Search Workflow" on page 91.		

Table 107. Workflow Editor menu (Sheet 2 of 2)

Command	Function	Shortcut	Toolbar icon
Workflow Editor > Auto Number	Automatically renumbers connected nodes in consecutive order in the workflow.		
Workflow Editor > Start Workflow	Begins the data search using the selected workflow. For information about starting a workflow, see "Creating a Search Workflow" on page 91.		
Workflow Editor > Import Workflow	Imports a workflow in XML or MSF format. For information about importing a workflow t, see "Opening an Existing Workflow from an XML or MSF File" on page 112.		
Workflow Editor > Export Workflow to XML	Exports a workflow to an XML-formatted file. See "Saving a Workflow as an XML Template" on page 114.		

Administration Menu

The Administration menu contains the following commands.

Table 108. Administration menu (Sheet 1 of 2)

Command	Function	Shortcut	Toolbar icon
Administration > Show Job Queue	Displays the job queue and the status of current searches. For more information about the job queue, see "Job Queue" on page 28	CTRL+J	
Administration > Maintain FASTA Files	Opens the FASTA files view on the Administration page, as shown in Figure 104 on page 160, so that you can add, remove, and modify FASTA files. For more information about FASTA files, see "Using FASTA Databases" on page 159.		
Administration > Maintain FASTA Indexes	Opens the FASTA files view on the Administration page, as shown in Figure 116 on page 182, so that you can add and remove FASTA indexes and change their location. For information about performing these functions, see "Managing FASTA Indexes" on page 179.		

Table 108. Administration menu (Sheet 2 of 2)

Command	Function	Shortcut	Toolbar icon
Administration > Maintain Chemical Modifications	Opens the Chemical Modifications view on the Administration page, as shown in Figure 124 on page 197, so you can enable or disable the available chemical modifications to be used in the search process. You can also create new modifications or edit existing modifications. For information about using the Chemical Modifications view, see "Chemical Modifications" on page 195.		
Administration > Maintain Cleavage Reagents	Opens the Cleavage Reagents view, as shown in Figure 130 on page 207, so that you can add, remove, and modify cleavage reagents. For more information about using cleavage reagents, see "Customizing Cleavage Reagents" on page 206.		
Administration > Maintain Quantification Methods	Opens the Quantification Methods view, as shown in Figure 267 on page 517, so that you can add, change, remove, import, or export a quantification method. For information about these functions, see "Adding a Quantification Method" on page 516, "Changing a Quantification Method" on page 520, "Removing a Quantification Method" on page 521, "Importing a Quantification Method" on page 522, and "Exporting a Quantification Method" on page 523.		
Administration > Manage Licenses	Opens the Licenses view in the Administration page so that you can add new licenses. For more information about adding new licenses or activating new features, see the *Proteome Discoverer Installation Guide*.		
Administration > Configuration	Opens the Configuration view in the Administration page so that you can configure the parameters used in all searches. For more information about configuring search parameters, see "Configuring Search Parameters" on page 64.		

Tools Menu

The Tools menu contains the following commands.

Table 109. Tools menu

Command	Function	Shortcut	Toolbar icon
Tools > Open QualBrowser	Starts Qual Browser, which enables you to view the entire ion chromatogram and browse individual precursor and MS^n data. You can filter the results in a variety of ways, for example, to produce a selected ion chromatogram. For more information about Qual Browser, see "Using the Qual Browser Application" on page 205. This command is only available if you click an individual peptide or row on the Search Input page to select it.	CTRL+SHIFT+B	
Tools > Export to ProteinCenter	Exports the search results in the MSF file to ProteinCenter. For information on this procedure, see "Uploading Results to ProteinCenter" on page 446.		
Tools > FASTA Database Utilities	Opens the FASTA Database Utilities dialog box, shown in Figure 106 on page 165. For information about the parameters in this dialog box, see "FASTA Database Utilities Dialog Box Parameters" on page 175.		
Tools > Create Bug Report	Opens the Proteome Discoverer Bug Reporter dialog box so you can enter a product change request. For information about this command, see the *Proteome Discoverer Installation Guide*.		
Tools > Mascot Communication Tester	Performs a series of tests to assist you in locating a problem if you cannot access the Mascot server. For information about this command, see the *Proteome Discoverer Installation Guide*.		
Tools > Options	Opens the Options dialog box, shown in Figure 25 on page 55. For information about the parameters in this dialog box, see "Options Dialog Box Parameters" on page 56.		

Window Menu

The Window menu contains the following commands.

Table 110. Window menu

Command	Function	Shortcut	Toolbar icon
Window > Close All Windows	Closes all the open windows in the Proteome Discoverer application session. It also lists any open pages, such as the Administration page or an MSF report page, so that you can close them.		

Help Menu

The Help menu contains the following commands.

Table 111. Help menu

Command	Function	Shortcut	Toolbar icon
Help > Help Contents	Displays the Help for the Proteome Discoverer application.	F1	
Help > Release Notes	Opens the *Proteome Discoverer Release Notes* in PDF format for the current release.		
Help > Installation Guide	Opens the *Proteome Discoverer Installation Guide* in PDF format for the current release.		
Help > User Guide	Opens the *Proteome Discoverer User Guide* in PDF format for the current release.		
Help > About Thermo Proteome Discoverer	Displays information about the installed version of the Proteome Discoverer application and its components. It also displays the Thermo Fisher Scientific copyright notice, patent and legal notices, and all license agreements.		

Shortcut Menus

The results report pages contain shortcut menus with corresponding commands.

- Proteins Page Shortcut Menu
- Peptides Page Shortcut Menu
- Search Input Page Shortcut Menu
- Graph Shortcut Menu
- Workflow Editor Shortcut Menu

Proteins Page Shortcut Menu

Table 112 lists the commands on the Proteins page shortcut menu.

Table 112. Proteins page shortcut menu (Sheet 1 of 3)

Command	Description
Show in Peptide View	Displays the Peptides page and highlights the peptide of interest. This command only appears on the shortcut menu when you click a peptide row on the Proteins page.
Show in Search Input View	Opens the Search Input page and highlights the peptide of interest. This command also displays any other peptides grouped with that scan number. This command only appears on the shortcut menu when you click a peptide row on the Proteins or Peptides page.
Show Peptide Groups	Displays as a group peptides that contain the same sequence.
Show Filtered Out Rows	Displays peptides and proteins that you have filtered out with the Enable Row Filters command or with the Result Filters in the shortcut menu. The filtered-out rows appear as two different shades of gray, depending on whether they were filtered by a row or a results filter.
Enable Protein Grouping	Determines whether all possible protein identifications are listed for the set of identified peptides, or whether they are grouped under the highest ranking protein (the master protein) to produce a non-redundant list.

Table 112. Proteins page shortcut menu (Sheet 2 of 3)

Command	Description
Mass Tolerance Unit	Specifies the units in which the values in the ΔM [ppm] column on the Peptides page are reported. This column displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion. • Da: Reports values in daltons. • ppm: Reports values in parts per million. • mmu: Reports values in milli-mass units.
Row Numbers	Determines whether the Proteome Discoverer application automatically assigns each row on the Proteins, Peptides, or Search Input page an index number. • None: Does not assign index numbers to any row. • Number only visible items: Adds row index numbers to visible rows only. • Number All Items (Including Currently Hidden Ones): Adds row index numbers to both visible and invisible rows. • Use Invariant Result Index: Adds fixed row index numbers to rows.
Copy	Copies the selected rows to the Clipboard. You can then paste the rows into a spreadsheet application such as Excel.
Copy With Column Headers	Copies the selected rows and the header captions to the Clipboard. You can then paste the rows into a spreadsheet application such as Excel.
Export to Excel Workbook	Copies and exports the specified proteins and the header captions to an Excel file.
Check Selected	Places a check mark next to selected proteins on the Proteins page.
Check All	Places a check mark next to all unfiltered proteins on the Proteins page.
Uncheck Selected	Removes a check mark next to selected proteins or peptides on the Proteins page.
Uncheck All	Removes a check mark next to all proteins or peptides the Proteins page.

Table 112. Proteins page shortcut menu (Sheet 3 of 3)

Command	Description
Enable Row Filters	Activates a logic operator and command menu icon,  , or a logic operator menu icon,  , in each column where you can access a menu to filter out types of data. Enable Row Filters also activates a down arrow,  , that opens a filter menu when clicked. For detailed information about row filters, see "Filtering Results with Row Filters" on page 393.
Show 'Group By Column' Panel	Enables grouping of data by column header.
Include/Exclude Peptide(s) From Protein Quantification	Determines whether the peptides are included or excluded from the quantification results. • Include: Includes the selected peptides in the quantification results and displays Used (Included) in the Quan Usage column. • Exclude: Excludes the selected peptides in the quantification results and displays Not Used (Excluded) in the Quan Usage column. • Default: Returns the selected peptides to the usage status they had when you first opened the MSF file and displays this status in the Quan Usage column. This command only appears on the shortcut menu when the MSF file includes quantification results.

Peptides Page Shortcut Menu

Table 113 lists the commands on the Peptides page shortcut menu.

Table 113. Peptides page shortcut menu (Sheet 1 of 3)

Parameter	Description
Show in Search Input View	Opens the Search Input page and highlights the peptide of interest. This command also displays any other peptides grouped with that scan number.
Show Peptide Groups	Displays as a group peptides that contain the same sequence.

Table 113. Peptides page shortcut menu (Sheet 2 of 3)

Parameter	Description
Show Filtered Out Rows	Displays peptides that you have filtered out from the Results Filters page or with the Result Filters command in the shortcut menu. The filtered-out rows appear as two different shades of gray, depending on whether they were filtered by a row or a results filter.
Enable Protein Grouping	Determines whether all possible protein identifications are listed for the set of identified peptides, or whether they are grouped under the highest ranking protein (the master protein) to produce a non-redundant list.
Mass Tolerance Unit	Specifies the units in which the values in the ΔM [ppm] column are reported. This column displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion. • Da: Reports values in daltons. • ppm: Reports values in parts per million. • mmu: Reports values in milli-mass units.
Row Numbers	Determines whether the Proteome Discoverer application automatically assigns each row on the Proteins, Peptides, or Search Input page an index number. • None: Does not assign index numbers to any row. • Number only visible items: Adds row index numbers to visible rows only. • Number All Items (Including Currently Hidden Ones): Adds row index numbers to both visible and invisible rows. • Use Invariant Result Index: Adds fixed row index numbers to rows.
Copy	Copies the selected rows to the Clipboard. You can then paste the rows in a spreadsheet application such as Excel.
Copy With Column Headers	Copies the selected rows and the header captions to the Clipboard. You can then paste the rows in a spreadsheet application such as Excel.
Export to Excel Workbook	Copies and exports the specified peptides and the header captions to an Excel file.

Table 113. Peptides page shortcut menu (Sheet 3 of 3)

Parameter	Description
Check Selected	Places a check mark next to selected proteins or peptides on the Peptides page.
Check All	Places a check mark next to all unfiltered proteins or peptides on the Peptides page.
Uncheck Selected	Removes a check mark next to selected proteins or peptides in the MSF report.
Uncheck All	Removes a check mark next to all proteins or peptides on the Peptides page.
Enable Row Filters	Activates a logic operator and command menu icon,  , or a logic operator menu icon,  , in each column where you can access a menu to filter out types of data. Enable Row Filters also activates a down arrow,  , that opens a filter menu when clicked. For detailed information about row filters, see "Filtering Results with Row Filters" on page 393.
Show 'Group By Column' Panel	Enables grouping of data by column header.
Include/Exclude Peptide(s) From Protein Quantification	Determines whether the peptides are included or excluded from the quantification results. • Include: Includes the selected peptides in the quantification results and displays Used (Included) in the Quan Usage column. • Exclude: Excludes the selected peptides in the quantification results and displays Not Used (Excluded) in the Quan Usage column. • Default: Returns the selected peptides to the usage status they had when you first opened the MSF file and displays this status in the Quan Usage column. This command only appears on the shortcut menu when the MSF file includes quantification results.

Search Input Page Shortcut Menu

Table 114 lists the commands on the Search Input page shortcut menu.

Table 114. Search Input page shortcut menu (Sheet 1 of 2)

Command	Description
Show Peptide Groups	Displays as a group peptides that contain the same sequence.
Show Filtered Out Rows	Displays peptides that you have filtered out with the Enable Row Filters command in the shortcut menu.
Enable Protein Grouping	Determines whether the most prevalent protein and its variations are grouped together.
Mass Tolerance Unit	Specifies the units in which the values in the ΔM [ppm] column on the Peptides page are reported. This column displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion. • Da: Reports values in daltons. • ppm: Reports values in parts per million. • mmu: Reports values in milli-mass units.
Row Numbers	Determines whether the Proteome Discoverer application automatically assigns each row on the Proteins, Peptides, or Search Input page an index number. • None: Does not assign index numbers to any row. • Number only visible items: Adds row index numbers to visible rows only. • Number All Items (Including Currently Hidden Ones): Adds row index numbers to both visible and invisible rows. • Use Invariant Result Index: Adds fixed row index numbers to rows.
Copy	Copies the selected rows to the Clipboard. You can then paste the rows into a spreadsheet application such as Excel.
Copy With Column Headers	Copies the selected rows and the header captions to the Clipboard. You can then paste the rows into a spreadsheet application such as Excel.

Table 114. Search Input page shortcut menu (Sheet 2 of 2)

Command	Description
Export to Excel Workbook	Copies and exports the specified search input rows and the header captions to an Excel file.
Check Selected	Places a check mark next to selected search input rows on the Search Input page.
Check All	Places a check mark next to all search input rows on the Search Input page.
Uncheck Selected	Removes a check mark next to selected rows on the Search Input page.
Uncheck All	Removes a check mark next to all rows on the Search Input page.
Enable Row Filters	Activates a logic operator and command menu icon,  , or a logic operator menu icon,  , in each column where you can access a menu to filter out types of data. Enable Row Filters also activates a down arrow,  , that opens a filter menu when clicked. For detailed information about row filters, see "Filtering Results with Row Filters" on page 393.
Show 'Group By Column' Panel	Groups peptide or protein data by column header.
Include/Exclude Peptide(s) From Protein Quantification	Determines whether the peptides are included or excluded from the quantification results. • Include: Includes the selected peptides in the quantification results and displays Used (Included) in the Quan Usage column. • Exclude: Excludes the selected peptides in the quantification results and displays Not Used (Excluded) in the Quan Usage column. • Default: Returns the selected peptides to the usage status they had when you first opened the MSF file and displays this status in the Quan Usage column. This command only appears on the shortcut menu when the MSF file includes quantification results.

Graph Shortcut Menu

You can right-click anywhere in any of the graphs displayed by the commands on the Search Report menu to display a shortcut menu. Not all commands on the shortcut menu apply to all graphs. Table 115 lists the commands on this shortcut menu.

Table 115. Graph shortcut menus

Command	Description
Show Position Tooltips	Displays information about a data point as a tooltip when you move your cursor over the data point in the plot.
Zoom Out	Zooms out to previous zoom perspective.
Undo All Zoom/Pan	Zooms out completely.
Copy	Copies the graph so that you can paste it into an Excel spreadsheet.
Save As	Opens the Save As dialog box so that you can save the graph in EMF, PNG, JPG, GIF, TIFF, or BMP format.
Show Legend	Displays a legend, if available, that explains components of the graph.

Workflow Editor Shortcut Menu

Table 116 lists the shortcut menu commands when you right-click in the workspace pane of the Workflow Editor.

Table 116. Workflow Editor shortcut menus

Command	Description
Cut (CTRL+X)	Copies the selected node to the Clipboard and deletes it from the current Workflow Editor view.
Copy (CTRL+C)	Copies the selected node to the Clipboard.
Paste (CTRL+V)	Inserts a copied node into the current Workflow Editor view.
Auto Layout	Automatically aligns the nodes in the workflow.
Auto Number	Automatically reorders the node numbering.

Workflow Editor Nodes

This section describes the nodes available in the Workflow Editor and the parameters that you can set for them. You can list the parameters for each node in alphabetical order by clicking on the alphabetical icon,  , in the Parameters pane of the Workflow Editor, or you can group the parameters into categories by clicking on the categorized icon,  , to the left of the alphabetical icon. You can display all the parameters available for a node by clicking Show Advanced Parameters.

This section lists the parameters for each node in alphabetical order.

For information about the “Is,” “Is Not,” and “Any” options offered by some of the node parameters, see “Creating a Search Workflow” on page 91.

- Data Input Nodes
- Spectrum & Feature Retrieval Nodes
- Spectrum Processing Nodes
- Spectrum Filters Nodes
- Peptide/Protein ID Nodes
- Peptide/Protein Validation Nodes
- PTM Analysis Nodes
- Quantification Nodes
- Data Export Nodes
- Annotation Nodes

Data Input Nodes

The Data Input nodes are available in the Data Input area of the Workflow Editor.

- Spectrum Files Node
- SRF File Selector Node

Spectrum Files Node

The Spectrum Files node specifies the name of the raw files to use for your search. You can select a single file, multiple files, or an entire folder. You cannot use the SRF File Selector node Selector with the Spectrum Files node.

Table 117. Spectrum Files node parameters

Parameters	Description
File name(s)	Opens the Select Analysis File(s) dialog box so that you can specify the name and location of the raw file or files or a folder of data.

SRF File Selector Node

The SRF File Selector node specifies the name of the SRF file to import. Use this node to open a BioWorks SRF file instead of a raw file for subsequent analysis or format conversion (see “Migrating from BioWorks to the Proteome Discoverer Application” on page 715).

Table 118. SRF File Selector node parameters

Parameters	Description
Absolute XCorr Threshold	Specifies a minimum cross-correlation threshold that determines whether peptides in an SRF file are imported. All peptides with an XCorr value less than this threshold are not imported. The default value is 0.4.
SRF File Name	Specifies the path and name of the SRF file to process.

Spectrum & Feature Retrieval Nodes

The Spectrum & Feature Retrieval nodes are available in the Spectrum & Feature Retrieval nodes area of the Workflow Editor.

- Event Detector Node
- Spectrum Selector Node

Event Detector Node

The Event Detector node clusters isotopes of precursor ions that elute during the same retention time. It also removes noise and spike signals from the spectra used for further processing. It is used in precursor ion quantification and peak area quantification.

Table 119. Event Detector node parameters

Parameters	Description
Mass Precision	Specifies the expected standard deviation of the mass precision. Three times the standard deviation is used to create extracted ion chromatograms. The minimum value is 0.5 ppm. The maximum value is 4 ppm. The default value is 2 ppm.
S/N Threshold	Specifies a threshold signal-to-noise value that determines whether the Proteome Discoverer application removes peaks from the spectrum. It removes peaks with a signal-to-noise value below this threshold. The minimum value is 1.0. There is no maximum value. The default value is 1.

Spectrum Selector Node

The Spectrum Selector node selects the part of the spectrum that you want to search. You must use this node with the Spectrum Files node.

When searches are based on MGF, MZDATA, and other files, the extracted spectra do not have information about the activation type, mass analyzer, polarity, and so forth, because this information is not necessarily available in these file formats. However, certain processing nodes require some of this information. For example, the ZCore node can only process ETD spectra. To enable such searches, you can use the Spectrum Selector node to set user-defined values for these spectrum properties in case this information is not available.

Note If you specify both a retention time range and a scan range, the Proteome Discoverer application combines them into a single filter criterion; that is, it processes a scan only if it simultaneously falls within the specified retention time range and the specified scan range.

Table 120. Spectrum Selector node parameters (Sheet 1 of 8)

Parameters	Description
Activation Type	Specifies the fragmentation method used to generate the MS/MS spectra. The following activation types are available: • CID (Collision-Induced Dissociation) • MPD (Multi-Photon Dissociation) • ECD (Electron Capture Dissociation) • PQD (Pulsed-Q Collision-Induced Dissociation) • ETD (Electron Transfer Dissociation) • HCD (High-Energy Collision Dissociation) • (Default) Any Activation Type For a description of these fragmentation types, see "Fragmentation Methods" on page 8.
First Scan	Specifies the scan number of the first scan to process. If you set this number to 0, the Proteome Discoverer application uses the first available scan. The minimum value is 0, and there is no maximum value. The default value is 0.

Table 120. Spectrum Selector node parameters (Sheet 2 of 8)

Parameters	Description
Highest Charge State	Filters out spectra with a precursor charge higher than the specified charge. If you set the value to zero, there is no maximum value for the precursor charge. The minimum value is 0. The default value is 0.
Ionization Source	Specifies the ionization source of the spectrum. The following sources are available: • Electrospray • Nanospray • Electron Impact • Atmospheric Pressure Chemical Ionization • Matrix Assisted Laser Desorption Ionization • Chemical Ionization • Fast Atom Bombardment • Field Desorption • Glow Discharge • (Default) Any
Last Scan	Specifies the scan number of the last scan to process. If you set this number to 0, the Proteome Discoverer application uses the last available scan. The minimum value is 0, and there is no maximum value. The default value is 0.

Table 120. Spectrum Selector node parameters (Sheet 3 of 8)

Parameters	Description
Lower RT Limit	Specifies the lower retention-time limit. Retention time is the time in the mass chromatogram when any particular precursor ion is observed. The Lower RT Limit and the Upper RT Limit parameters set limits on the portion of the data that you want to analyze. Since very early and very late portions in the chromatogram typically do not contain the peptide ions of interest, you might want to analyze only that middle portion that contains useful data to save search time. If the value is 0.0, the lowest available retention time is used. The minimum value is 0. There is no maximum value. The default value is 0.
Lowest Charge State	Filters out spectra with a precursor charge lower than the specified charge. The minimum value is 0. There is no maximum value. The default value is 0.
Mass Analyzer	Specifies the type of mass analyzer used to acquire the spectrum. This parameter filters out all but the type you want. It is possible to have data from more than one type of mass analyzer in the same data input, for example, data from a hybrid instrument such as LTQ Orbitrap or LTQ FT. The available analyzers are the following: • Ion Trap • Fourier Transform (FT) (Fourier transform applies to both FT ICR data and Orbitrap data.) • Time of Flight (TOF) • Single Quad • Triple Quad • Sector Field • (Default) Any

Table 120. Spectrum Selector node parameters (Sheet 4 of 8)

Parameters	Description
Max. Precursor Mass	Specifies the upper mass limit of a singly charged precursor ion to be processed. If you set the value to zero, the value is set to the maximum precursor mass allowed. The minimum value is 0.0 Da. The maximum value is 10000.0 Da. The default value is 5000 Da.
Min. Precursor Mass	Specifies the lower mass limit of a singly charged precursor ion to be processed. The minimum value is 0.0 Da. The maximum value is 10000.0 Da. The default value is 350 Da.
Minimum Peak Count	Specifies the minimum number of peaks in a tandem mass spectrum that is allowed to pass the filter and to be subjected to further processing in the workflow. Use this parameter to remove spectra that have a small number of peaks, because these typically produce low-quality hits. You can use any value. The minimum value is 1, and there is no maximum value. The default value is 1.
MS Order	Specifies the level of the mass spectrum to be processed, for example, MS/MS or MS^3. You can choose from MS^1 through MS^{10}. The default value is MS2.
Polarity Mode	Specifies the polarity mode of the spectrum. You can select one of the following settings: • Positive: The detector monitors positive ions. • Negative: The detector monitors negative ions. • (Default) Any: The detector monitors either positive or negative ions.

Table 120. Spectrum Selector node parameters (Sheet 5 of 8)

Parameters	Description
Precursor Selection	Determines which precursor mass to use for a given MS^n scan. This option applies only to higher-order MS^n scans like MS^3, MS^4, and so forth. • (Default) Use MS1 Precursor: Uses the precursor mass in the MS scan. • Use MS(n – 1) Precursor: Uses the precursor mass in the parent scan. For example, for MS^3, this option would use the MS/MS scan.
S/N Threshold (FT-only)	Specifies a signal-to-noise value below which peaks are removed. The S/N threshold has an effect on FT data only. Data acquired with other mass analyzers is not affected. The minimum value is 0.0, and there is no maximum value. The default value is 3.
Scan Type	Specifies the scan type of the spectrum. The available scan types are the following: • (Default) Full: Scans the full mass range to acquire the data. The mass spectrometer detects a wide range of mass fragments. This option is used for typical data-dependent scanning. • Selected Ion Monitoring (SIM): Scans a very small mass range to acquire the data. The mass spectrometer detects only certain mass precursors. • Single Reaction Monitoring (SRM): The mass spectrometer detects specific product ions from specific parent ions.
Total Intensity Threshold	Specifies a minimum total intensity value and filters out tandem mass spectra that have a total intensity current (the sum of the intensities of all peaks in a spectrum) below the specified value. You can use this parameter to remove low-intensity data. You can use any number. The minimum value is 0.0, and there is no maximum value. The default value is 0.

Table 120. Spectrum Selector node parameters (Sheet 6 of 8)

Parameters	Description
Unrecognized Activation Type Replacements	Specifies the fragmentation method to use in the search algorithm if it is not included in the scan header. You can select from the following methods: • (Default) CID (Collision-Induced Dissociation) • MPD (Multi-Photon Dissociation) • ECD (Electron Capture Dissociation) • PQD (Pulsed-Q Collision-Induced Dissociation) • ETD (Electron Transfer Dissociation) • HCD (High-Energy Collision Dissociation) For a description of these fragmentation types, see "Fragmentation Methods" on page 8.
Unrecognized Charge Replacements	Specifies the charge state of the ions to be included in the search, if it is not in the scan header. If the charge state of the precursor cannot be determined, the spectrum is replicated for each of the specified precursor charges. You can select from the following settings: • (Default) Automatic: Automatically assigns the +2 and +3 charge states to any spectrum whose precursor charge state is not defined. • 1 – 8: Assigns a charge state that you select from +1 to +8, inclusive, to any spectrum whose precursor charge state is not defined. If you select Automatic, the Proteome Discoverer application applies a simple check for a +1 spectrum; otherwise, it generates a +2 and a +3 spectrum.

Table 120. Spectrum Selector node parameters (Sheet 7 of 8)

Parameters	Description
Unrecognized Mass Analyzer Replacements	Specifies the mass spectrometer to use to produce the spectra, if it is not included in the scan header. You can select from the following settings: • (Default) Ion Trap (ITMS) • Fourier Transform (FTMS) • Time of Flight (TOFMS) • Single Quad (SMS) • Triple Quad (TMS) • Sector Field (SectorMS)
Unrecognized MS Order Replacements	Specifies the MS scan order used to produce the product spectra, if it is not included in the scan header. You can select from MS1 through MS10. The default is MS2.
Unrecognized Polarity Replacements	Specifies the polarity of the ions monitored, if it is not included in the scan header. You can select from the following settings: • (Default) Positive (+): Specifies that the detector monitored positive ions. • Negative (–): Specifies that the detector monitored negative ions.

Table 120. Spectrum Selector node parameters (Sheet 8 of 8)

Parameters	Description
Upper RT Limit	Specifies the upper retention-time limit. If the value is 0.0, the highest available retention time is used. The minimum value is 0, and there is no maximum value. The default value is 0.
Use New Precursor Reevaluation	Determines the type of precursor reevaluation algorithm used to detect the correct isotopic pattern of the precursor. • (Default) True: Uses a sophisticated algorithm to obtain the best precursor peak for a given MS/MS spectrum from the corresponding master scan. The MS1 master scans must have been acquired in the FT instrument. Using the monoisotopic mass and charge reported by the instrument in the spectrum header (or if that is not available, the isolation mass), the algorithm goes into the master scan and tries to find the corresponding peak in the MS1 spectrum both to improve the mass accuracy of the monoisotopic mass reported by the instrument and to correct for inaccurate precursor masses corresponding to one of the ^{13}C isotopes. • False: Uses a simplified algorithm that looks for matching masses only to the left of a reported peak and performs a rough check of the detected isotopic intensities.

Spectrum Processing Nodes

The Spectrum Processing nodes are available in the Spectrum Processing area of the Workflow Editor.

- ETD Spectrum Charger Node
- Noise Peak Filter Node
- Non-Fragment Filter Node
- Spectrum Grouper Node

- Spectrum Normalizer Node
- Top N Peaks Filter Node

ETD Spectrum Charger Node

The ETD Spectrum Charger node calculates precursor ion charge states for ETD spectra when the charge state of the precursor is not defined in the scan header. This node has no parameters. This node works on both low-resolution and high-resolution MS/MS spectra but is only needed if the MS1 scan is acquired on a low-resolution detector like an LTQ.

Each MS/MS spectrum has a precursor from an MS1 spectrum. With hybrid instruments such as an LTQ Orbitrap or LTQ FT, the following combinations of MS/MS and related MS1 scans can result:

- MS1 (FT detector)/MS/MS (FT detector)
- MS1 (FT detector)/MS/MS (ion trap detector)
- MS1 (ion trap detector)/MS/MS (FT detector)
- MS1 (ion trap detector)/MS/MS (ion trap detector)

If the MS1 scans are acquired with an FT detector, as in the first two cases just listed, the charge state of the precursor is already determined by the FT instrument, so there is no need to determine the charge state again with the ETD Spectrum Charger node.

If the MS1 scans are acquired with an ion trap detector, the charge state of the precursor is not determined by the instrument and must be determined when processing the data. In this case, you can use the ETD Spectrum Charger node. In principle, this mode (the third scenario in the list just given) is not used with a hybrid instrument, because the mass of the precursor is determined as accurately as possible.

The last mode given in the list is also not used on a hybrid instrument, but it is the normal operation mode of an ion-trap-only instrument, such as an LTQ.

Noise Peak Filter Node

The Noise Peak Filter node removes all peaks from the processed mass lists that fall below the specified signal-to-noise ratio threshold. The Noise Peak Filter works only on mass lists derived from spectra acquired with an FT mass analyzer. It has no effect if used with spectra acquired with other mass analyzers, such as ion traps. This node is useful if you want to divide your workflow into two or more paths and want to assign different signal-to-noise thresholds to each path.

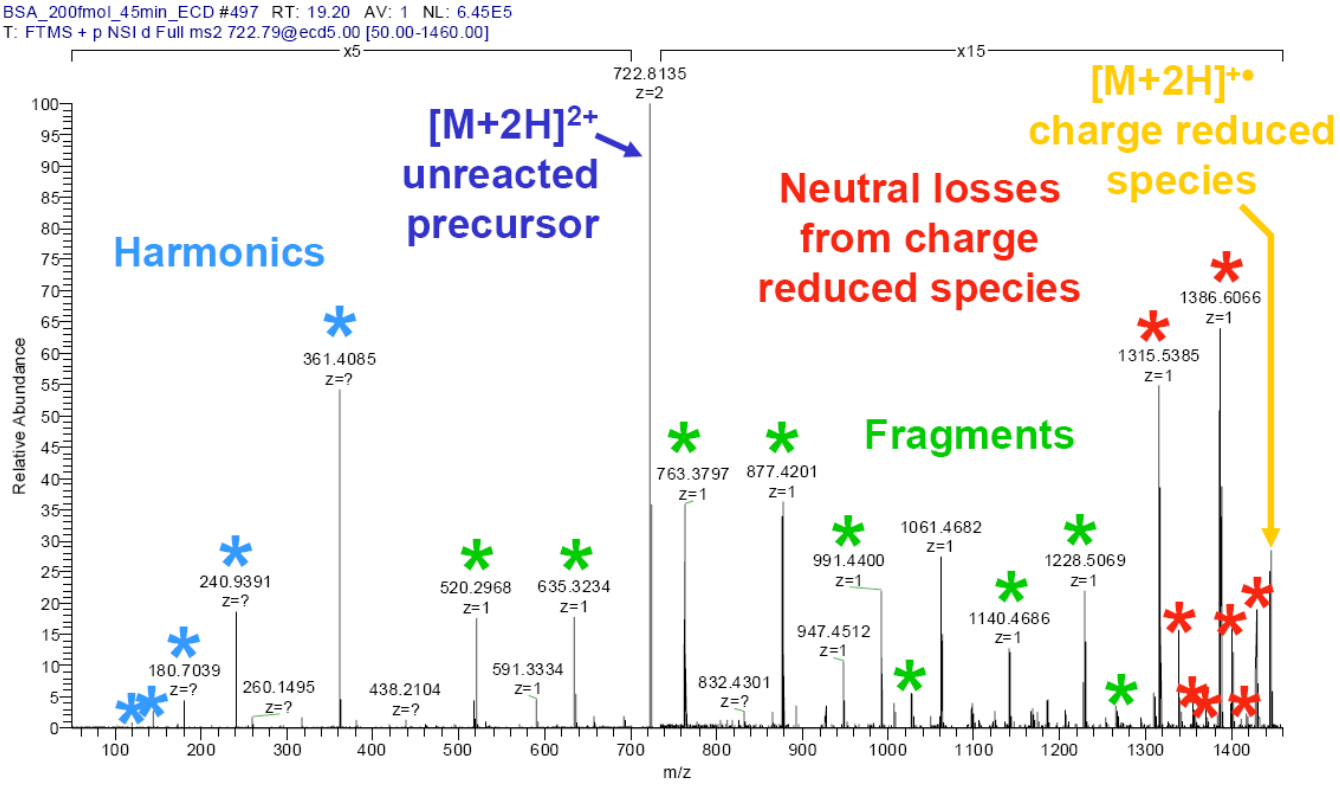
Table 121. Noise Peak Filter node parameters

Parameters	Description
S/N Threshold (FT-only)	Specifies a signal-to-noise value below which peaks are removed. The S/N threshold has an effect on FT data only. Data acquired with other mass analyzers is not affected. The minimum value is 0.0, and there is no maximum value. The default value is 3.

Non-Fragment Filter Node

The Non-Fragment Filter node removes precursor artifacts from the MS/MS input spectra. Examples of these artifacts are the peaks of unfragmented precursor ions, peaks of charged-reduced species of precursor ions and their resulting neutral loss peaks, and precursor artifact peaks due to higher harmonic frequencies (overtones). ETD and ECD fragment spectra typically include two types of ion peaks: true fragment ion peaks and peaks that are related to the precursor ions. Figure 333 gives an example of a spectrum that contains these two types of ion peaks. The Non-Fragment Filter node calculates and removes non-fragment ion peaks from the mass lists that are generated before submission to database searches. Non-fragment ion peaks can lead to false positive identifications in database searches that use search algorithms that score the experimentally generated spectrum versus a calculated theoretical spectrum. The removed peaks are not likely to be fragment peaks from ETD or ECD data. In the following example, this node is restricted to precursor ion peaks present in ETD and ECD spectra.

Figure 333. Example of a typical ECD spectrum (BSA, peptide YICDNDTISSK)



For information about adding a non-fragment filter node for Orbitrap or FT ICR data, see "Creating a Search Workflow for Multiple Raw Files from the Same Sample" on page 101.

Given the mass of the precursor, you can calculate the theoretical mass-to-charge ratio (*m/z*) values of these non-fragment peaks and remove every peak within these regions with a certain mass tolerance.

Calculate these mass windows as follows:

- Precursor peaks: The final mass window used to remove precursor peaks is the following:

  window_lower_mass = precursor_m/z - offset

  window_upper_mass = precursor_m/z + (#isotope_peaks/precursor_charge) + offset

  The Proteome Discoverer application looks up the number of isotope peaks (*#isotope_peaks*) in the static table (shown in Table 122) according to the uncharged mass.

Table 122. Uncharged mass and number of isotope peaks used in calculating precursor peaks

Uncharged mass	Isotope peaks
< 1000	3
1000 – 2000	4
2000 – 3000	5

Table 122. Uncharged mass and number of isotope peaks used in calculating precursor peaks

Uncharged mass	Isotope peaks
3000 – 4000	6
4000 – 5000	7
5000 – 6000	8
6000 – 7000	9
> 7000	10

- Precursor overtone peaks: The final mass window used to remove precursor overtone peaks is as follows:

  window_lower_mass = *overtone_m/z* - *offset*

  window_upper_mass = *overtone_m/z* + (*#isotope_peaks*/*precursor_charge*) + *offset*

  where the *overtone_m/z* is at 1/2, 1/3, 1/4 of the precursor mass-to-charge ratio (*m/z*).

- Charge-reduced precursor peaks: The final mass window used to remove charge-reduced precursor peaks is as follows:

  window_lower_mass = *charge_reduced_precursor_m/z* - (*proton_mass*/*charge*) - *offset*

  window_upper_mass = *charge_reduced_precursor_m/z* + (*#isotope_peaks*/*charge*) + *offset*

  where *charge* is the charge of the reduced species. *Charge_reduced_precursor_m/z* is calculated by adding electrons to the precursor until charge 1 is reached but without removing protons. For example, if the precursor ion is [M+4H]4+, the next charge-reduced ion is [M+4H]3+, then [M+4H]2+, and so forth.

- Neutral loss peaks of charge-reduced precursor ions: The final mass window used to remove neutral loss peaks of charge-reduced precursor ions depends on whether all peaks within a large neutral loss mass window should be removed or only the peaks within smaller windows around the positions of known neutral losses from charge-reduced species.

  When you remove only small windows around known neutral losses, calculate the used mass windows as follows:

  window_lower_mass = *neutral_loss_m/z* - *offset*

  window_upper_mass = *neutral_loss_m/z* + (*#isotope_peaks*/*charge*) + *offset*

  where *charge* is the charge of the reduced species.

  If you must remove a large mass range to eliminate neutral loss peaks, calculate the used mass window as follows:

  window_lower_mass = *charge_reduced_precursor_m/z* - (*maximum_loss_mass*/ *charge*)

  window_upper_mass = *charge_reduced_precursor_m/z*

If you must remove only known neutral loss masses, use the values in Table 123.

Table 123. Neutral loss masses

Mass	Neutral loss
17.027 Da	NH_3
18.011 Da	H_20
27.995 Da	CO
32.026 Da	CH_3OH
34.053 Da	N_2H_6 ($2xNH_3$)
35.037 Da	H_4NO
36.021 Da	H_4O_2 ($2xH_2O$)
44.037 Da	CH_4N_2
45.021 Da	CH_3NO
46.006 Da	CH_2O_2
46.042 Da	C_2H_6O
59.037 Da	C_2H_5NO
59.048 Da	CH_5N_3
73.089 Da	$C_4H_{11}N$
74.019 Da	C_3H_6S
82.053 Da	$C_4H_6N_2$
86.072 Da	$C_3H_8N_3$
99.068 Da	$C_4H_9N_3$
101.095 Da	$C_4H_{11}N_3$
108.058 Da	C_7H_8O
131.074 Da	C_9H_9N

Table 124 describes the parameters available for the Non-Fragment Filter node.

Table 124. Non-Fragment Filter node parameters (Sheet 1 of 3)

Parameters	Description
Mass Window Offset	Specifies the size of the mass-to-charge ratio (*m/z*) window, in daltons, from which the Remove Charge Reduced Precursors option removes charge-reduced precursor peaks. The minimum value is 0.0 Da, and there is no maximum value. The default value is 0.5 Da.
Mass Window Offset	Specifies the tolerance for the Remove Neutral Loss Peaks option, in daltons. The minimum value is 0.0 Da, and there is no maximum value. The default value is 0.5 Da.
Mass Window Offset	Specifies the size of the mass-to-charge ratio (*m/z*) window, in daltons, from which the Remove Precursor Overtones option removes precursor overtone peaks. The minimum value is 0.0 Da, and there is no maximum value. The default value is 0.5 Da.
Mass Window Offset	Specifies the size of the mass-to-charge ratio (*m/z*) window, in daltons, from which the Remove Precursor Peak option removes precursor artifact peaks from the spectrum. The minimum value is 0.0 Da, and there is no maximum value. The default value is 1 Da.
Maximum Neutral Loss Mass	Specifies the maximum allowed neutral loss. The minimum value is 0.0 Da, and there is no maximum value. The default value is 120 Da.

Table 124. Non-Fragment Filter node parameters (Sheet 2 of 3)

Parameters	Description
Remove Charge Reduced Precursors	Determines whether the Proteome Discoverer application removes the charge-reduced precursor peaks found in an spectrum. These precursors only complicate the analysis of the fragment spectra. The Proteome Discoverer application removes all peaks within a given mass window around the precursor mass-to-charge ratio. • (Default) True: Removes the charge-reduced precursor peaks found in a spectrum. • False: Does not remove the charge-reduced precursor peaks found in a spectrum.
Remove Neutral Loss Peaks	Determines whether the Proteome Discoverer application removes neutral loss peaks from the spectra. Charge-reduced species formed in reactions can undergo neutral loss reactions, creating non-fragment peaks that complicate the analysis by typical search engines. The Proteome Discoverer application removes all peaks within a given mass window around the precursor mass-to-charge ratio. • (Default) True: Removes neutral loss peaks from the spectra. • False: Does not remove neutral loss peaks from the spectra.
Remove Only Known Masses	Determines whether the Proteome Discoverer application removes overtone peaks from LTQ FT or LTQ FT Ultra spectra. The Proteome Discoverer application removes all peaks within a given mass window around the precursor mass-to-charge ratio. • (Default) True: Removes overtone peaks from LTQ FT or LTQ FT Ultra spectra. • False: Does not remove overtone peaks from LTQ FT or LTQ FT Ultra spectra.

Table 124. Non-Fragment Filter node parameters (Sheet 3 of 3)

Parameters	Description
Remove Precursor Overtones	Determines whether the Proteome Discoverer application removes precursor overtone peaks in the spectrum from the input spectrum. Overtones, or harmonics, are an artifact of FT detectors. The Proteome Discoverer application removes all peaks within a given mass window around the precursor mass-to-charge ratio. • (Default) True: Removes precursor overtone peaks from the input spectrum. • False: Does not remove precursor overtone peaks from the input spectrum.
Remove Precursor Peak	Determines whether the Proteome Discoverer application removes precursor artifact peaks from the MS/MS input spectra. The Proteome Discoverer application removes all peaks within a given mass window around the precursor mass-to-charge ratio. • (Default) True: Removes precursor artifact peaks from the MS/MS input spectra. • False: Does not remove precursor artifact peaks from the MS/MS input spectra.

Spectrum Grouper Node

The Spectrum Grouper node applies a grouping function to the data set. You can use this node to group spectra from different MS order scans or different mass analyzers.

The Spectrum Grouper node combines the MS/MS peak lists that belong to the same precursor species within specified tolerances. This feature can lead to a reduction in the number of peak lists, faster search times, and a higher quality of the individual peak lists.

The Spectrum Grouper steps through the peak lists that it receives and combines them according to defined criteria. The result of this grouping is a combined fragment spectrum peak list that is associated with a single precursor mass and charge. The peak lists are then used as the input for the peptide identification search.

You must be precise when defining the grouping parameters. Setting the tolerances too wide leads to combining fragment spectra of multiple or unique species. This type of setting decreases the quality of the search results and can result in fewer peptide hits, more false positives, or overall lower probability scores. Setting the grouping tolerances too tight can result in longer search times and a possible decrease in the quality of data, because species that could be combined are treated as unique. In some cases, it may be beneficial to group spectra if a low-abundance peptide produces a poor-quality (low signal-to-noise) spectrum.

You can also use the Spectrum Grouper node to group HCD and CID scans to improve the overall fragment ion coverage in alternating HCD/CID experiments. For this latter purpose, set the Allow MS Order Mismatch parameter to True. The default is False.

The Spectrum Grouper node cannot group spectra from multiple input files.

You can use the Spectrum Grouper node either through the Workflow Editor or through the search wizards.

Table 125 describes the parameters available for the Spectrum Grouper node.

Table 125. Spectrum Grouper node parameters (Sheet 1 of 2)

Parameter	Description
Allow Mass Analyzer Mismatch	Determines whether the fragment spectrum for scans with the same precursor mass is grouped, regardless of mass analyzer and activation type. • True: Sorts the fragment spectrum for scans with the same precursor mass into the same group, regardless of mass analyzer and activation type. For example, both ITMS and FTMS MS/MS scans are sorted into the same group, as long as they have the same precursor mass. • (Default) False: Does not group the fragment spectrum for scans with the same precursor mass if they arise from different mass analyzers. These scans are analyzed separately. **Note**: When you attempt to group ITMS and FTMS scans, also use the Spectrum Normalizer node to average these two types of data since they are of very different intensities. The Spectrum Normalizer node normalizes for effective averaging. For information about the Spectrum Normalizer node, see "Spectrum Normalizer Node" on page 658.
Allow MS Order Mismatch	Determines whether spectra from different MS order scans (for example, MS2 and MS3) can be grouped together. This option might be useful, for example, for analyzing MS^2 + MS^3 data from a data dependent neutral loss experiment. • True: Groups spectra from different MS order scans together. • (Default) False: Does not group spectra from different MS order scans together.

Table 125. Spectrum Grouper node parameters (Sheet 2 of 2)

Parameter	Description
Max. RT Difference (min)	Specifies the chromatographic window where precursors to be grouped must reside to be considered the same species and therefore added to the same group. For example, a precursor ion with a mass-to-charge ratio of 619 that elutes at 37.76 minutes is different from a precursor ion with a mass-to-charge ratio of 619 that elutes at 47.10 minutes. The minimum value is 0.0, and there is no maximum value. The default value is 1.5 minutes.
Precursor Mass Criterion	Groups spectra measured within the given mass and retention-time tolerances into a single spectrum for analysis on the basis of a predesignated precursor mass criterion. You can select one of the following settings: • (Default) Same Measured Mass-to-Charge: Averages the tandem mass spectrum of precursors that have the same mass-to-charge value within the specified tolerance. • Same Singly Charged Mass: Groups different charge states of the same precursor together. It averages the tandem mass spectrum of precursors that have the same singly charged mass within a specified tolerance. This option is only applicable when the charge state of the precursor is recorded in the original data, as would be the case with higher-resolution data such as FT ICR or Orbitrap data. The charge associated with the combined peak list is the highest charge of the precursors in the group.
Precursor Mass Tolerance	Specifies the mass window for which precursor ions are considered to be the same species and are therefore added to the same group. They must also elute within the specified mass tolerance window. The minimum value is 0.0 Da or 0.0 ppm. The maximum value is 3.5 Da or 5000 ppm. The default value is 10 ppm.

Spectrum Normalizer Node

The Spectrum Normalizer node normalizes the intensities of a spectrum so that the largest peak has an arbitrary intensity of 100. This node is useful when you combine two or more scans through the Spectrum Grouper node, and one of the original spectra has a very different amplitude than the other. For example, if you are trying to group ITMS and FTMS scans, use the Spectrum Normalizer node to average the intensity of these very different data types.

This node has no parameters.

Top *N* Peaks Filter Node

The Top *N* Peaks Filter node improves the number of peptides identified with high confidence by reducing the number of peaks in the searched peak lists. It avoids matching peptide candidates to spurious or noise peaks, so it produces no false peptide matches.

A commonly used method when searching proteomics data is to remove all but the *n* most abundant peaks within a given mass window from the generated peak lists. Normal settings keep only the top six to ten peaks within every 100 Da mass window.

You can place the Top *N* Peaks Filter node anywhere between the Spectrum Selector node and a search engine node, such as Mascot or SEQUEST. You can use multiple Top *N* Peaks Filter nodes within the same workflow, if necessary.

Table 126 describes the parameters available for the Top *N* Peaks Filter node.

Table 126. Top N Peaks Filter node parameters

Parameter	Description
Top *N*	Specifies the number of the most abundant peaks to keep within the specified mass window. The default value is 6. The minimum value is 1. There is no maximum value.
Mass Window	Specifies the size of the mass window, in daltons. The Proteome Discoverer application divides the spectrum into mass windows of the specified size, and in each of these mass windows, it filters out all peaks except the number of the most abundant peaks specified by the Top *N* parameter. The default value is 100 Da. The minimum value is 1.0 Da. There is no maximum value.

The Top *N* Filter node divides each input spectrum or peak list into consecutive mass windows of the specified size. The first mass window is placed at *m/z* = 0 Da. After that, this filter removes all peaks except the specified number of the most abundant peaks from every mass window.

You might prefer to filter the top *n* most abundant peaks within the entire spectrum, for example, the top 50 peaks of every spectrum. Set the Mass Window parameter to a large width that always spans the whole mass range of the entire spectrum, for example, 10 000 Da. Set the Top *N* parameter to the number of the most abundant peaks to keep for every spectrum, for example, 50.

Spectrum Filters Nodes

The Spectrum Filters nodes are available in the Spectrum Filters area of the Workflow Editor.

- Scan Event Filter Node
- Spectrum Properties Filter Node
- Spectrum Confidence Filter Node

Scan Event Filter Node

The Scan Event Filter node enables you to divide data from a "mixed-mode" experiment so that each "mode" can be processed independently. For example, if you have a raw file obtained from alternating ETD and CID scans, you might want to analyze the CID portion of the data with SEQUEST and the ETD portion with ZCore.

Table 127 describes the parameters available for the Scan Event Filter node.

Table 127. Scan Event Filter node parameters (Sheet 1 of 3)

Parameter	Definition
Activation Type	Specifies the fragmentation method used to activate the scan: • CID (Collision-Induced Dissociation) • MPD (Multi-Photon Dissociation) • ECD (Electron Capture Dissociation) • PQD (Pulsed Collision-Induced Dissociation) • ETD (Electron Transfer Dissociation) • HCD (High-Energy Collision Dissociation) • (Default) Any Activation Type For a description of these fragmentation types, see "Fragmentation Methods" on page 8.
Ionization Source	Specifies the ionization source of the spectrum. The following sources are available: • Electrospray • Nanospray • Thermospray • Electron Impact • Atmospheric Pressure Chemical Ionization • Matrix Assisted Laser Desorption • Chemical Ionization • Fast Atom Bombardment • Field Desorption • Glow Discharge • (Default) Any

Table 127. Scan Event Filter node parameters (Sheet 2 of 3)

Parameter	Definition
Mass Analyzer	Specifies the type of mass analyzer used for the acquisition of the spectrum. The following analyzers are available: • Ion Trap (ITMS) • Fourier Transform (FTMS) • Time of Flight (TOFMS) • Single Quad (SQMS) • Triple Quad (TQMS) • Sector Field (SectorMS) • (Default) Any
MS Order	Specifies the level of the mass spectrum to be processed, for example, MS/MS or MS^3. You can choose from MS1 through MS10. The default is Any.

Table 127. Scan Event Filter node parameters (Sheet 3 of 3)

Parameter	Definition
Polarity Mode	Specifies the polarity mode of the spectrum. You can select one of the following settings: • Positive: The detector monitors positive ions. • Negative: The detector monitors negative ions. • (Default) Any: The detector monitors both positive and negative ions.
Scan Type	Specifies the scan type to be used for the search. The available scan types are the following: • Full: Scans the full mass range to acquire the data. The mass spectrometer detects a wide range of mass fragments. This option is used for typical data-dependent scanning. • Selected Ion Monitoring (SIM): Scans a very small mass range to acquire the data. The mass spectrometer detects only certain mass fragments. • Single Reaction Monitoring (SRM): Scans a single mass in the first analyzer and a single user-defined fragment ion in the second analyzer to acquire the data. The mass spectrometer detects a specific product ion of a specific parent ion. • (Default) Any: Uses any of the three scan types just described.

Spectrum Properties Filter Node

The Spectrum Properties Filter node sets thresholds for the data that you pick to analyze. It performs the same function as the Spectrum Selector node but is included separately so that you can use separate thresholds in different branches of a branched workflow.

Table 128 describes the parameters available for the Spectrum Properties Filter node.

Note If you specify both a retention time range and a scan range, the Proteome Discoverer application combines them into a single filter criterion; that is, it processes a scan only if it simultaneously falls within the specified retention time range and the specified scan range.

Table 128. Spectrum Properties Filter node parameters (Sheet 1 of 2)

Parameter	Definition
First Scan	Specifies the scan number of the first scan to process. If you set this number to 0, the Proteome Discoverer application uses the first available scan. The minimum value is 0, and there is no maximum value. The default value is 0.
Highest Charge State	Specifies the maximum charge above which peptides are filtered out. The minimum value is 0, and there is no maximum value. The default value is 0.
Last Scan	Specifies the scan number of the last scan to process. If you set this number to 0, the Proteome Discoverer application uses the last available scan. The minimum value is 0, and there is no maximum value. The default value is 0.
Lower RT Limit	Specifies the lower retention-time limit. If you set this value to 0.0, the Proteome Discoverer application uses the lowest available retention time. The minimum value is 0, and there is no maximum value. The default value is 0.
Lowest Charge State	Specifies the minimum precursor charge state below which spectra are filtered out. The minimum value is 0, and there is no maximum value. The default value is 0.
Max. Precursor Mass	Specifies the maximum mass limit of a singly charged precursor ion to be processed, in daltons. If you set the value to zero, the Proteome Discoverer application applies no upper limit to the precursor mass. The minimum value is 0, and there is no maximum value. The default value is 0 Da.

Table 128. Spectrum Properties Filter node parameters (Sheet 2 of 2)

Parameter	Definition
Min. Precursor Mass	Specifies the minimum mass limit of a singly charged precursor ion to be processed, in daltons. The minimum value is 0, and there is no maximum value. The default value is 0 Da.
Minimum Peak Count	Specifies the minimum number of peaks that a tandem spectrum must have in order to pass the filter. The minimum value is 1, and there is no maximum value. The default value is 1.
Total Intensity Threshold	Specifies the minimum total intensity that a tandem mass spectrum must have in order to pass the filter. The minimum value is 0, and there is no maximum value. The default value is 0.
Upper RT Limit	Specifies the upper retention-time limit. If you set this value to 0.0, the Proteome Discoverer application uses the highest available retention time. The minimum value is 0, and there is no maximum value. The default value is 0.

Spectrum Confidence Filter Node

The Spectrum Confidence Filter node sets up iterative searches. You can specify confidence levels for the purpose of re-searching all spectra that do not have a peptide identified with this confidence level. In the second search, you might apply wider tolerances, search against a bigger taxonomy, use other modifications, and so forth. The second search can therefore evaluate data that failed to meet the confidence level of the first search by using a different set of parameters—for example, evaluation with PTMs not included in the first search.

Table 129. Spectrum Confidence Filter node parameters

Parameter	Definition
Spectrum Confidence	Specifies a confidence criterion and filters out spectra that have peptide matches that do not meet this criterion. • Equals: Filters out spectra that have peptide matches that are not of the specified confidence level. • Better Than: Filters out spectra that have peptide matches that are of a higher confidence level than the specified level. • Worse Than: Filters out spectra that have peptide matches that are of a lower confidence level than the specified level. • Low: Indicates a low-confidence peptide. • Medium: Indicates a medium-confidence peptide. • High: Indicates a high-confidence peptide. Here are some examples: • If you set the Spectrum Confidence parameter to Worse Than High, the Proteome Discoverer application passes all spectra that do not have a high-confidence peptide, including spectra with no peptides. • If you set the Spectrum Confidence parameter to Equals Medium, the application passes spectra having medium-confidence peptides. • If you set the Spectrum Confidence parameter to Worse Than Low, the application passes spectra without identified peptides. • If you set the Spectrum Confidence parameter to Better Than High, the application filters out all spectra.

Peptide/Protein ID Nodes

The Peptide/Protein ID nodes are available in the Peptide/Protein ID area of the Workflow Editor.

- Mascot Node
- SEQUEST Node
- ZCore Node

Mascot Node

The Mascot node analyzes the data with the Mascot search engine.

Table 130 describes the parameters available for the Mascot node.

Table 130. Mascot node parameters (Sheet 1 of 5)

Parameter	Definition
Dynamic Modifications	Determine PTMs. Some amino acid residues are modified and some are not modified.
Enzyme Name	Specifies the enzyme reagent used for protein digestion. The following enzymes are typical of what might be available. They are set up by the Mascot server. • Arg-C • Asp-N • Asp-N_ambic • Chymotrypsin_loose • CNBr • CNBr+Trypsin • Formic_acid • Lys-C • Lys-C/P • LysC+AspN • None • PepsinA • SemiTrypsin • TrypChymo • Tryp-CNBr • (Default) Trypsin • Trypsin/P • V8-DE • V8-E

Table 130. Mascot node parameters (Sheet 2 of 5)

Parameter	Definition
Fragment Mass Tolerance	Specifies the mass tolerance used for matching fragment peaks. You can also specify the units (Da or mmu) for the fragment mass tolerance value. The minimum value is 0.0001 Da, and the maximum value is 2.0 Da. The default value is 0.8 Da.
Instrument	Specifies the type of instrument used to acquire the data in the raw file. You can select from among the following: • ESI-4SECTOR • ESI-FTICR • ESI-QUAD • ESI-QUAD-TOF • ESI-TRAP • ETD-TRAP • ETD-TRAP-A-IONS • FTMS-ECD • HCD test • MALDI Orbitrap • MALDI-QIT-TOF • MALDI-QUAD-TOF • MALDI-TOF-PSD • MALDI-TOF-TOF • Orbi-HCD-DMH
Maximum Missed Cleavage Sites	Specifies the maximum number of missed cleavage sites to consider during the digest. The minimum value is 0, and the maximum value is 9. The default value is 1.

Table 130. Mascot node parameters (Sheet 3 of 5)

Parameter	Definition
Peptide Cut Off Score	Sets a minimum score in the IonScore column that each peptide must exceed in order to be reported. If a peptide has a score less than this minimum, it is discarded. The minimum value is 0, and there is no maximum value. The default value is 10.
Peptide Without Protein Cut Off Score	Specifies a threshold that determines whether the Proteome Discoverer application stores peptides that do not belong to a protein. It stores peptides with an IonScore value above this threshold and discards peptides with an IonScore value below this threshold. The minimum value is 0, and there is no maximum value. The default value is 5.
Precursor Mass Tolerance	Specifies the mass tolerance used for finding peptide candidates. You can also specify the units (Da, mmu, or ppm) for the precursor mass tolerance value. The minimum value is 0.0001 Da or 0.04 ppm. The maximum value is 5.0 Da or 5000 ppm. The default value is 10 ppm.
Protein Database	Specifies the database to use in the search. The available databases are configured on the Mascot server.
Protein Relevance Factor	Specifies a factor that is used in calculating a threshold that determines whether a protein appears in the results report. A protein must include at least one peptide having a value in the IonScore column above the product of *factor* × the homology threshold (or at least 13) in order to appear in the report. The minimum value is 0.0, and the maximum value is 2.0. The default value is 1.

Table 130. Mascot node parameters (Sheet 4 of 5)

Parameter	Definition
Protein Relevance Threshold	Specifies a threshold that determines whether a protein appears in the results report. A protein must have at least one peptide appear in the results report with a value in the Ions Score column above this threshold. The minimum value is 0, and there is no maximum value. The default value is 20.
Static Modifications	Apply the same specific mass to all occurrences of that named amino acid, as in an exhaustive chemical modification.
Taxonomy	Limits searches to entries from a particular species or group of species. This parameter can speed up a search and ensure that the hit list will only contain entries from the selected species. Select from the taxonomy database list.

Table 130. Mascot node parameters (Sheet 5 of 5)

Parameter	Definition
Use Average Precursor Mass	Determines the type of mass to use for matching the precursor. • True: Uses the average precursor mass. • (Default) False: Uses the monoisotopic mass.
Use MudPIT Scoring	Determines whether to use MudPIT scoring or normal scoring. MudPIT scoring is an option for Mascot searches that is used for generating the protein score from data sets with large numbers of peptide matches, as in a complex proteome digest. Mascot stores 10 possible matches for each spectrum. The protein score is normally a sum of all possible matches. However, this might result in a high "protein" score that is composed of numerous low-scoring erroneous matches. The Use MudPIT Scoring option prevents low-scoring peptide matches from contributing to a falsely reported high-scoring protein. • (Default) Automatic: Automatically determines whether to use MudPIT scoring. • True: Always uses MudPIT scoring. • False: Never uses MudPIT scoring. For information about MudPIT scoring, consult the Mascot Help.

SEQUEST Node

The SEQUEST node analyzes the data with the SEQUEST search engine.

Table 131 describes the parameters available for the SEQUEST node.

Table 131. SEQUEST node parameters (Sheet 1 of 7)

Parameter	Definition
Absolute XCorr Threshold	Specifies a threshold below which peptides are discarded from the report. The minimum value is 0.0, and there is no maximum value. The default value is 0.4.
Calculate Probability Scores	Determines whether the Proteome Discoverer application calculates a probability score for every peptide match. This score is an assessment of the probability that the reported match is a random occurrence. A lower probability score indicates a better match. • True: The application calculates the probability score. • (Default) False: The application does not calculate the probability score.
C-Terminal Modification	Specifies the dynamic C-terminal modification that is used during the search. The default value is None.
Dynamic Modifications	Determine PTMs. Some amino acid residues are modified and some are not modified. The available modifications are defined in the Chemical Modifications view on the Administration page, which you open by choosing Administration > Maintain Chemical Modifications.

Table 131. SEQUEST node parameters (Sheet 2 of 7)

Parameter	Definition
Enzyme Name	Specifies the enzyme reagent used for protein digestion. After you select your enzyme, you can also specify the type of enzymatic cleavage to consider (fully enzymatic, partially enzymatic, limited to C terminal, or limited to N terminal). To add to the list of enzymes, see "Adding a Cleavage Reagent" on page 211. The available enzymes are the following: • (Default) Trypsin • Trypsin (KRLNH) • Chymotrypsin • Chymotrypsin (FWY) • Clostripain • Cyanogen_Bromide • IodosoBenzoate • Proline Endopept • Staph_Protease • Trypsin_K • Trypsin_R • GluC • LysC • AspN • Elastase • Elastase/Tryp/Chymotrypsin • No-Enzyme
Fragment Ion Cutoff Percentage	Specifies the percentage of the theoretical ions that must be found in order for a peptide to be scored and retained. The minimum value is 0.0, and the maximum value is 1.0. The default value is 0.1.
Fragment Mass Tolerance	Specifies the mass tolerance used for matching fragment peaks. You can also specify the units (Da or mmu) for the fragment mass tolerance value. The minimum value is 0.0001 Da, and the maximum value is 2.0 Da. The default value is 0.8 Da.

Table 131. SEQUEST node parameters (Sheet 3 of 7)

Parameter	Definition
Max. Modifications Per Peptide	Specifies the maximum number of different modifications that a peptide can have. For example, it is not feasible for a modification to exist on every amino acid of a peptide because there would be too much steric hindrance. The minimum value is 0, and the maximum value is 12. The default value is 4.
Maximum Missed Cleavage Sites	Specifies the maximum number of missed cleavage sites to consider during the digest. The minimum value is 0, and the maximum value is 12. The default value is 2.
Maximum Peptides Considered	Specifies the maximum number of peptides that are searched and scored per spectrum. The minimum value is 1, and the maximum value is 1000. The default value is 500.
Maximum Peptides Output	Specifies the maximum number of peptide matches reported per spectrum. The minimum value is 1, and the maximum value is 500. The default value is 10.
Maximum Protein References Per Peptide	Specifies the maximum number of proteins that a single identified peptide can be associated with. This number affects the ultimate protein list assembled from the set of identified peptides. The minimum value is 1, and the maximum value is 100. The default value is 100.
N-Terminal Modification	Specifies the dynamic N-terminal modification that is used during the search. The default value is None.
Peptide C-Terminus	Specifies the static modification for the C terminal of the peptide used during the search. The default value is None.

Table 131. SEQUEST node parameters (Sheet 4 of 7)

Parameter	Definition
Peptide N-Terminus	Specifies the static modification for the N terminal of the peptide used during the search. The default value is None.
Peptide Relevance Factor	Specifies a factor to apply to the protein score. The Proteome Discoverer application calculates the protein score for a protein as follows: *protein_score* = (*sum_of_all_cross-correlation_factors_of_0.8_or_above*) + (*peptide_charge* × *peptide_relevance_factor*) The minimum value is 0.0, and the maximum value is 0.8. The default value is 0.4.
Peptide Without Protein XCorr Threshold	Specifies a threshold that determines whether the Proteome Discoverer application stores peptides that do not belong to a protein. It stores peptides with an XCorr value above this threshold and discards peptides with an XCorr value below this threshold. The minimum value is 0, and there is no maximum value. The default is 1.5
Precursor Mass Tolerance	Specifies the mass tolerance used for finding peptide candidates. You can also specify the units (Da, mmu, or ppm) for the precursor mass tolerance value. The minimum value is 0.0001 Da or 0.01 ppm. The maximum value is 5.0 Da or 5000 ppm. The default value is 10 ppm.
Protein Database	Specifies the database to use in the search. You determine the list of available FASTA files.

Table 131. SEQUEST node parameters (Sheet 5 of 7)

Parameter	Definition
Protein Relevance Threshold	Specifies a peptide threshold that determines whether the protein that it is a part of is scored and retained in the report. The protein score is a composite of supporting peptide identification XCorr values. If the protein has at least one peptide above the specified threshold, the protein is scored and retained. The minimum value is 0.0, and the maximum value is 10.0. The default value is 1.5.
Static Modifications	Apply the same specific mass to all occurrences of that named amino acid, as in an exhaustive chemical modification. The available modifications are defined in the Chemical Modifications view on the Administration page, which you open by choosing Administration > Maintain Chemical Modifications.
Use Average Fragment Masses	Determines the type of mass to use for matching the fragment. • True: Uses the average mass. • (Default) False: Uses the monoisotopic mass.
Use Average Precursor Mass	Determines the type of mass to use for matching the precursor. • True: Uses the average mass. • (Default) False: Uses the monoisotopic mass.
Use Neutral Loss a Ions	Determines whether a ions with neutral loss are used for spectrum matching. • (Default) True: Uses a ions with neutral loss. • False: Does not use a ions with neutral loss.
Use Neutral Loss b Ions	Determines whether b ions with neutral loss are used for spectrum matching. • (Default) True: Uses b ions with neutral loss. • False: Does not use b ions with neutral loss.

Table 131. SEQUEST node parameters (Sheet 6 of 7)

Parameter	Definition
Use Neutral Loss y Ions	Determines whether y ions with neutral loss are used for spectrum matching. • (Default) True: Uses y ions with neutral loss. • False: Does not use y ions with neutral loss.
Weight of a Ions	Uses a ions for spectrum matching with this relative factor. The minimum value is 0.0, and the maximum value is 1.0. The default value is 0.
Weight of b Ions	Uses b ions for spectrum matching with this relative factor. The minimum value is 0.0, and the maximum value is 1.0. The default value is 1.
Weight of c Ions	Uses c ions for spectrum matching with this relative factor. The minimum value is 0.0, and the maximum value is 1.0. The default value is 0.
Weight of x Ions	Uses x ions for spectrum matching with this relative factor. The minimum value is 0.0, and the maximum value is 1.0. The default value is 0.

Table 131. SEQUEST node parameters (Sheet 7 of 7)

Parameter	Definition
Weight of y Ions	Uses y ions for spectrum matching with this relative factor. The minimum value is 0.0, and the maximum value is 1.0. The default value is 1.
Weight of z Ions	Uses z ions for spectrum matching with this relative factor. The minimum value is 0.0, and the maximum value is 1.0. The default value is 0.

ZCore Node

The ZCore node analyzes the data with the ZCore search engine. ZCore can only process ETD spectra.

Table 132 describes the parameters available for the ZCore node.

Table 132. ZCore node parameters (Sheet 1 of 3)

Parameter	Definition
Dynamic Modifications	Determine PTMs. Some amino acid residues are modified and some are not modified. The available modifications are defined in the Chemical Modifications view on the Administration page, which you open by choosing Administration > Maintain Chemical Modifications.
Enzyme Name	Specifies the enzyme reagent used for protein digestion. After you select your enzyme, you can also specify the type of enzymatic cleavage to consider (fully enzymatic, partially enzymatic, limited to C terminal, or limited to N terminal). To add to the list of enzymes, see "Adding a Cleavage Reagent" on page 211. The available enzymes are the following: • (Default) Trypsin • Trypsin (KRLNH) • Chymotrypsin • Chymotrypsin (FWY) • Clostripain • Cyanogen_Bromide • IodosoBenzoate • Proline Endopept • Staph_Protease • Trypsin_K • Trypsin_R • GluC • LysC • AspN • Elastase • Elastase/Tryp/Chymotrypsin • No-Enzyme
Max. Modifications Per Peptide	Specifies the maximum number of different modifications that a peptide can have. For example, it is not feasible for a modification to exist on every amino acid of a peptide because there would be too much steric hindrance. The minimum value is 0, and the maximum value is 12. The default value is 4.

Table 132. ZCore node parameters (Sheet 2 of 3)

Parameter	Definition
Maximum Missed Cleavage Sites	Specifies the maximum number of missed cleavage sites to consider during the digest. The minimum value is 0, and the maximum value is 12. The default value is 2.
Maximum Peptides Output	Specifies the maximum number of peptides reported per spectrum. The minimum value is 1, and the maximum value is 500. The default is 10.
Maximum Protein References Per Peptide	Specifies the maximum number of proteins that a single identified peptide can be associated with. This number affects the ultimate protein list assembled from the set of identified peptides. The minimum value is 1, and the maximum value is 100. The default is 100.
Peptide C-terminus	Specifies the chemical modification for the C terminal of the peptide used during the search. The default value is None.
Peptide N-terminus	Specifies the chemical modification for the N terminal of the peptide used during the search. The default value is None.
Peptide Without Protein Score Cut Off	Specifies a threshold that determines whether the Proteome Discoverer application stores peptides that do not belong to a protein. It stores peptides with an Probability value above this threshold and discards peptides with an Probability value below this threshold. The minimum value is 0, and there is no maximum value. The default is 15.

Table 132. ZCore node parameters (Sheet 3 of 3)

Parameter	Definition
Precursor Mass Tolerance	Specifies the mass tolerance used for finding peptide candidates. You can also specify the units (Da, mmu, or ppm) for the precursor mass tolerance value. The minimum value is 0.0001 Da or 0.01 ppm, and the maximum value is 5.0b Da or 5000 ppm. The default value is 1 Da.
Protein Database	Specifies the database to use in the search. You determine the list of the available FASTA files.
Protein Score Cutoff	Sets a minimum protein score that each protein must exceed in order to be reported. This score is a composite of the supporting peptide identification scores. If a protein has a score less than this minimum, it is discarded. The minimum value is 0, and there is no maximum value. The default value is 30.
Static Modifications	Apply the same specific mass to all occurrences of that named amino acid, as in an exhaustive chemical modification. The available modifications are defined in the Chemical Modifications view on the Administration page, which you open by choosing Administration > Maintain Chemical Modifications.

Peptide/Protein Validation Nodes

The nodes in the Peptide/Protein Validation section of the Workflow Nodes pane calculate the confidence of peptide identification in the search results.

- Peptide Validator Node
- Percolator Node

Peptide Validator Node

The Peptide Validator node calculates the probability that the search algorithm incorrectly included a peptide in a sample.

Table 133 describes the parameters available for the Peptide Validator node.

Table 133. Peptide Validator node parameters

Parameter	Definition
Search Against Decoy Database	Determines whether the Proteome Discoverer application searches a decoy database, which is a search against both true and reversed database sequences, in order to determine the false discovery rate for each given dataset. • (Default) True: Searches the decoy database. • False: Does not search the decoy database.
Target FDR (Strict)	Specifies the strict target false discovery rate for peptide hits with high confidence. The minimum value is 0.0, and the maximum value is 1.0. The default is 0.01.
Target FDR (Relaxed)	Specifies the relaxed target false discovery rate for peptide hits with moderate confidence. The minimum value is 0.0, and the maximum value is 1.0. The default is 0.05.

Percolator Node

To estimate the number of false positive protein identifications in a more systematic fashion, you can use a decoy database containing reversed protein sequences. Since this initial application, many other researchers have used decoy searches to estimate the number of incorrect PSMs that exceed a given threshold. This approach allows you to adjust the score threshold to obtain a target false discovery rate[1,2].

Although database-searching algorithms work well, current scoring methods produce significant overlap between the scores of correct and incorrect peptide identifications. To ensure that a large fraction of the true positive identifications are retained, you must select a score threshold to ensure that a percentage of the peptide identifications is incorrect.

Because most database search algorithms return multiple scores—for example, XCorr, Sp, and Delta Cn for SEQUEST—most proteomics studies apply separate thresholds to each score. Using multiple orthogonal score criteria is useful for eliminating false discoveries that might exceed one threshold but not another. However, in most cases these orthogonal scores are considered independently, ignoring the benefits that can be obtained if the features are considered jointly[3].

[1] Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*, **2007**, *4*: 207–214.

[2] Käll, L.; Storey, J. D.; MacCoss, M. J. & Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res*, **2008**, *7*: 29–34.

The Percolator node uses semi-supervised machine learning to discriminate correct from incorrect peptide spectrum matches and calculates accurate statistics, such as the q-value (FDR) and posterior error probability (PEP), to improve the number of confidently identified peptides at a given false discovery rate. It also assigns a statistically meaningful q-value to each PSM and the probability of the individual PSM being incorrect[4].

Methodology

The Percolator algorithm uses a semi-supervised method to train a machine learning algorithm called a support vector machine (SVM) to discriminate between positive and negative PSMs[3,5,6]. As negative examples for the classifier, Percolator uses the PSMs derived from searching a decoy database that consists of reversed protein sequences (the classifier is the feature vector with its determined weights). As positive examples, it uses a subset of the high-scoring PSMs derived from searching the target database.

For each target and decoy PSM, Percolator computes a set of features related to the quality of the match—for example, search engine scores, precursor mass deviation, and average fragment mass deviation. Subsequently, Percolator chooses the most relevant feature from the set of target and decoy PSMs and filters it to a fixed false discovery rate (FDR)—for example, 1 percent. It then applies the learned classifier to all target and decoy PSMs and again filters the most relevant feature to a fixed FDR to continue the training procedure. After a few iterations, the system converges and results in a robust classifier that then rescores each PSM in the data set. The whole process is fully automated and does not require any expert-driven or subjective decisions, eliminating any artificial biases. The learned classifier is specifically adapted to each data set and is unique to it, allowing adaptation to variations in data quality, protocols, and instrumentation.

Note To work properly, Percolator needs a sufficient number of PSMs from the target and the decoy search. If the search identified fewer than 200 target or decoy PSMs, or if fewer than 20 percent decoy PSMs are available compared to the number of target matches, Percolator rejects them for processing and displays an appropriate message in the Proteome Discoverer job queue or in the Search Summary of an open report. In these cases, open the MSF file when the search finishes and perform a target- and decoy-based FDR calculation in the Peptide Confidence page.

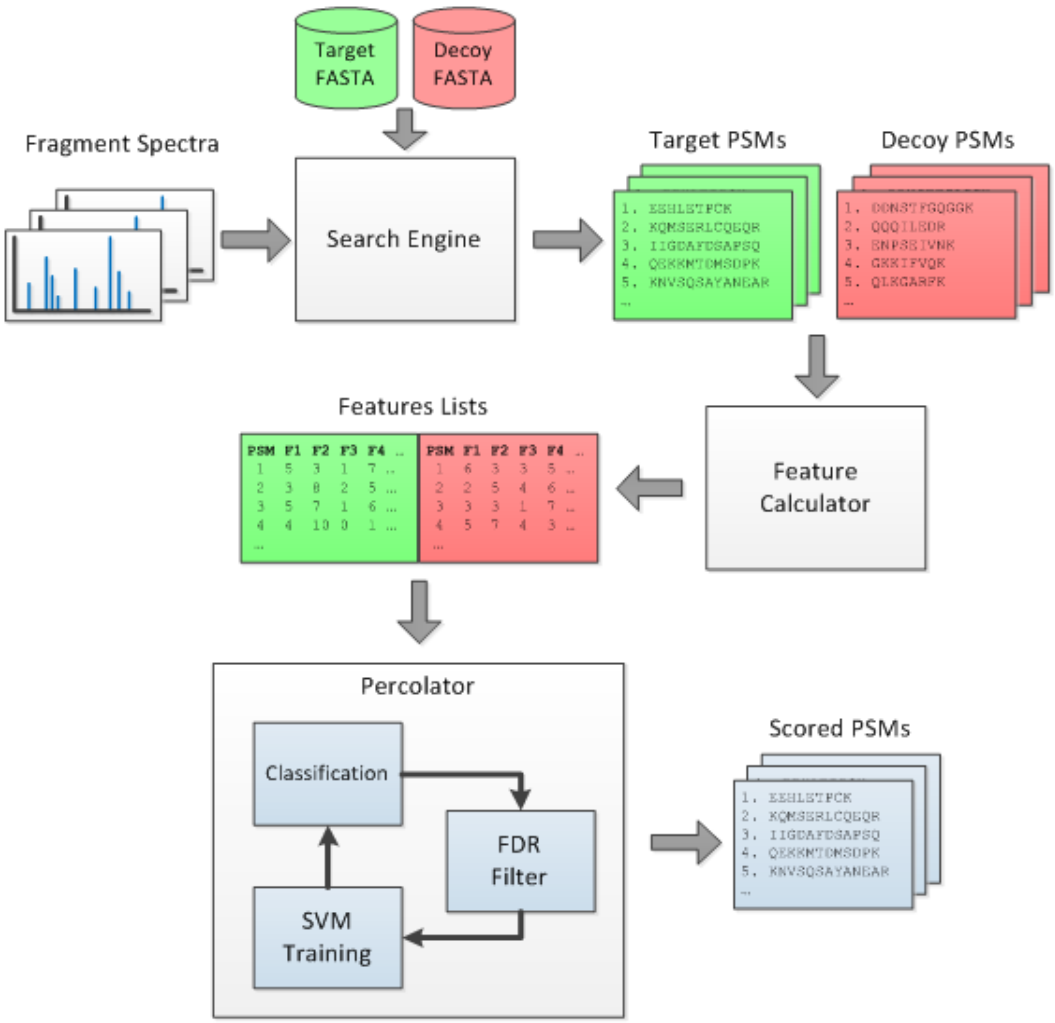
[3] Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*, **2007** *4*: 923–925.

[4] Käll, L.; Storey, J. D.; MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res*, **2008**, *7*: 40–44.

[5] Brosch, M.; Yu, L.; Hubbard, T. & Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. *J Proteome Res*, **2009**, *8*: 3176–3181.

[6] Spivak, M.; Weston, J.; Bottou, L.; Käll, L. & Noble, W. S. Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. *J Proteome Res*, **2009**, *8*: 3737–3745.

Figure 334. Assigning q-values and PEPs to a set of target and decoy PSMs generated by a database search engine



Feature Sets

Except for the individual scores of the search engines, the Proteome Discoverer application uses the same feature set for all available search engines. Table 134 lists the features in this feature set.

Table 134. Features used for Percolator in the Proteome Discoverer application (Sheet 1 of 4)

Feature	Description
Search engine/scoring-specific features	
Mascot Ions Score	Refer to the Mascot documentation from Matrix Science. This feature is only used for Mascot searches.

Table 134. Features used for Percolator in the Proteome Discoverer application (Sheet 2 of 4)

Feature	Description
SEQUEST Xcorr	Scores the number of fragment ions that are common to two different peptides with the same precursor mass and calculates the cross-correlation score for all candidate peptides queried from the database (SEQUEST searches only).
SEQUEST Sp Score	Only used for SEQUEST searches.
Delta Cn	Normalized score distance from the second-best-scoring PSM of the spectrum.
Binomial Score	Binomial score as described by S. A. Beausoleil, J. Villen, S.A. Gerber, J. Rush, and S.P. Gygi.[a]
Peptide-precursor-related features	
% Isolation Interference	Fraction of ion current in the isolation window not attributed to the identified precursor.
MH+ [Da]	Singly charged mass of the peptide.
Delta Mass [Da]	Deviation of the measured mass from the theoretical mass of the peptide, in daltons.
Delta Mass [ppm]	Deviation of the measured mass from the theoretical mass of the peptide, in ppm.
Absolute Delta Mass [Da]	Absolute deviation of the measured mass from the theoretical mass of the peptide, in daltons.
Absolute Delta Mass [ppm]	Absolute deviation of the measured mass from the theoretical mass of the peptide, in ppm.
Peptide Length	Length of the peptide in residues.
Is z=1	Binary flag indicating whether the charge state of the PSM is 1.
Is z=2	Binary flag indicating whether the charge state of the PSM is 2.
Is z=3	Binary flag indicating whether the charge state of the PSM is 3.
Is z=4	Binary flag indicating whether the charge state of the PSM is 4.

Table 134. Features used for Percolator in the Proteome Discoverer application (Sheet 3 of 4)

Feature	Description
Is z=5	Binary flag indicating whether the charge state of the PSM is 5.
Is z>5	Binary flag indicating whether the charge state of the PSM is above 5.
Digestion-related features	
# Missed Cleavages	Number of missed cleavages.
FASTA-related features	
Log Peptides Matched	Logarithm of the number of candidates in the precursor mass window.
Spectrum-related features	
Log Total Intensity	Logarithm of the total ion current of the fragment spectrum.
Fraction Matched Intensity [%]	Fraction of the total ion current of the fragment spectrum that is matched by fragments of the PSM.
Fragment-series-related features	
Fragment Coverage Series A, B, C [%]	Coverage of the N-terminal fragment ion series. The coverage is separately calculated for each series used, and the maximum coverage is used.
Fragment Coverage Series X, Y, Z [%]	Coverage of the C-terminal fragment ion series. The coverage is separately calculated for each series used, and the maximum coverage is used.
Log Matched Fragment Series Intensities A, B, C	Logarithm of the intensity sum of all matched N-terminal fragment ion series peaks. The intensity sum is separately calculated for each series used, and the maximum coverage is used.
Log Matched Fragment Series Intensities X, Y, Z	Logarithm of the intensity sum of all matched C-terminal fragment ion series peaks. The intensity sum is separately calculated for each series used, and the maximum coverage is used.

Table 134. Features used for Percolator in the Proteome Discoverer application (Sheet 4 of 4)

Feature	Description
Longest Sequence Series A, B, C	Longest consecutive matched sequence among the N-terminal fragment ion series peaks. The sequence length is separately calculated for each series used, and the maximum coverage is used.
Longest Sequence Series X, Y, Z	Longest consecutive matched sequence among the C-terminal fragment ion series peaks. The sequence length is separately calculated for each series used, and the maximum coverage is used.
Fragment-related features	
IQR Fragment Delta Mass [Da]	Inter-quartile range of the distribution of mass errors of all fragments considered, in daltons.
IQR Fragment Delta Mass [ppm]	Inter-quartile range of the distribution of mass errors of all fragments considered in ppm.
Mean Fragment Delta Mass [Da]	Arithmetic mean range of the distribution of mass errors of all fragments considered in daltons.
Mean Fragment Delta Mass [ppm]	Arithmetic mean range of the distribution of mass errors of all fragments considered in ppm.
Mean Absolute Fragment Delta Mass [Da]	Arithmetic mean range of the distribution of absolute mass errors of all fragments considered, in daltons.
Mean Absolute Fragment Delta Mass [ppm]	Arithmetic mean range of the distribution of absolute mass errors of all fragments considered, in ppm.

[a] Beausoleil, S.A.; Villen, J.; Gerber, S.A.; Rush, J. and Gygi, S.P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*, **2006**, *24:* 1285–1292.

The Search Summary lists the weights that Percolator assigns to individual features, as shown in Figure 335. The job queue also displays these weights, as Figure 336 demonstrates.

Figure 335. Weights that Percolator assigns to individual features shown in the Search Summary

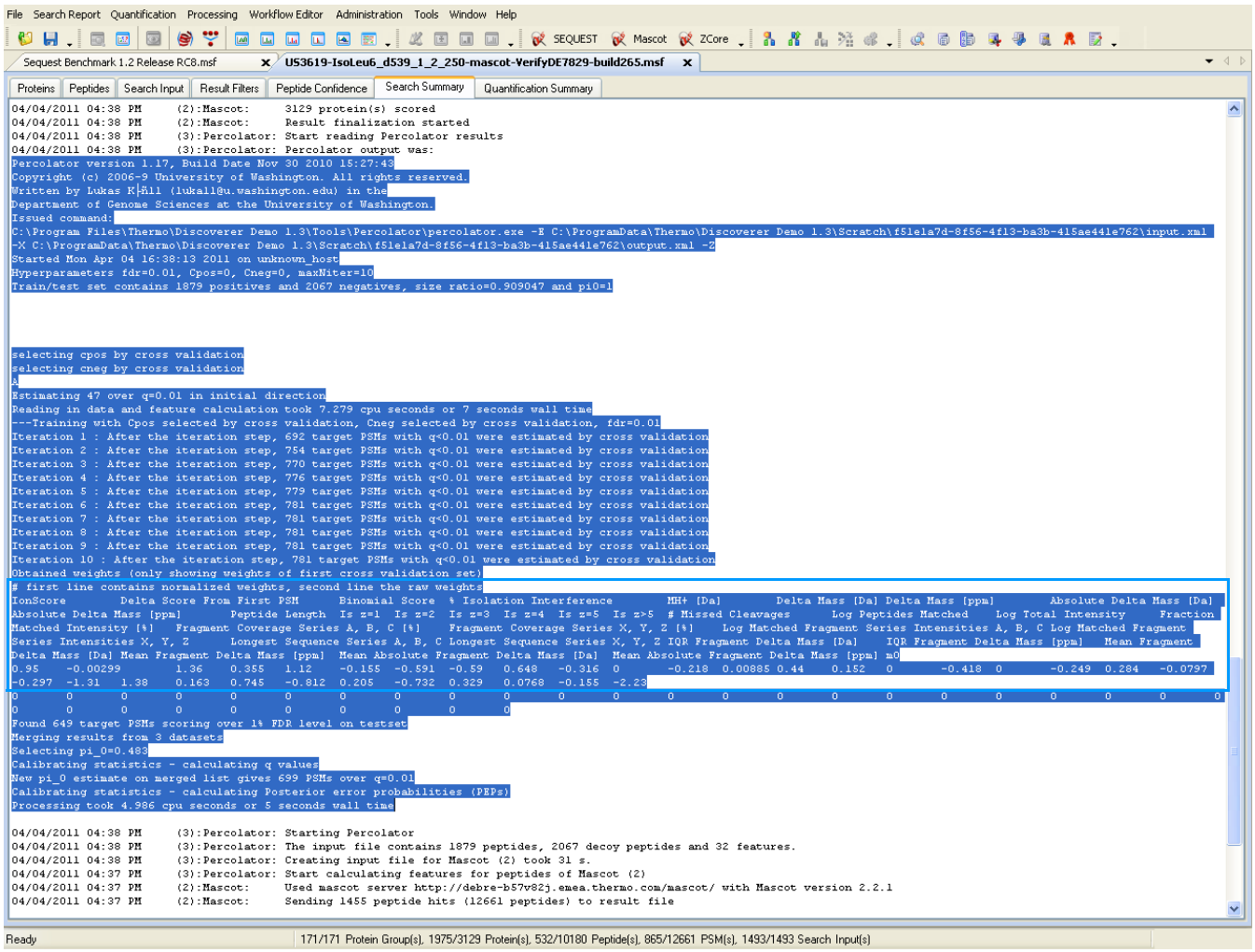
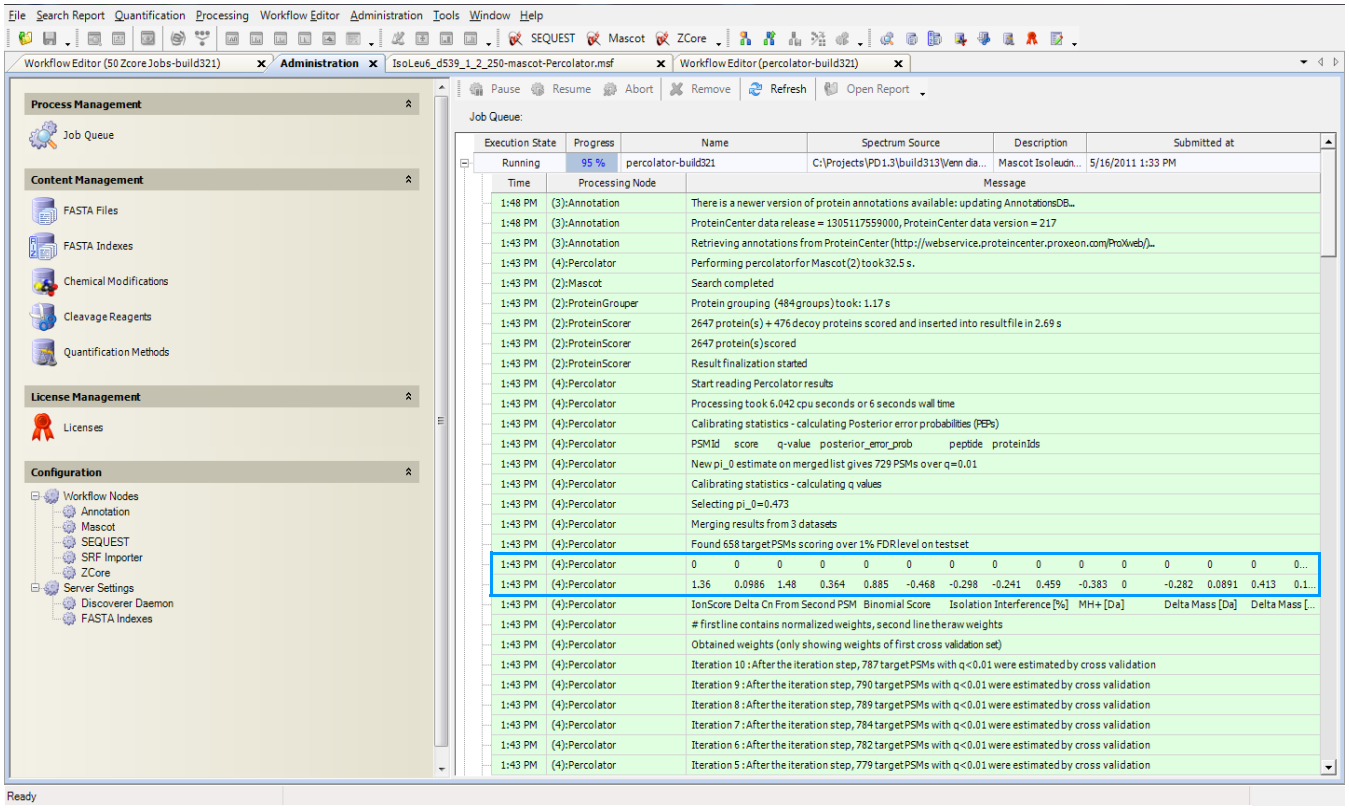


Figure 336. Weights that Percolator assigns to individual features shown in the job queue



q-value

A q-value is the minimal false discovery rate at which the identification is considered correct[7]. q-values are estimated using the distribution of scores from the decoy database search. A q-value of 0.01 for the EAMRQPK peptide matching spectrum, *s*, means that if you try all possible FDR thresholds, 1 percent is the minimal FDR threshold at which the PSM of EAMRQPK to *s* appears in the output list.

Although the q-value is associated with a single PSM, it also depends on the data set that the PSM occurs in.

Posterior Error Probability (PEP)

The posterior error probability (PEP) is the probability that the observed PSM is incorrect. For example, if the PEP associated with (EAMRPK, *s*) is 5 percent, there is a 95 percent chance that the EAMRPK peptide was in the mass spectrometer when spectrum *s* was generated.

[7] Käll, L.; Storey, J. D.; MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res*, **2008**, *7*: 40–44.

The PEP is similar to a local version of the FDR. The FDR measures the error rate associated with a collection of PSMs, and the PEP measures the probability of error for a single PSM. Similarly, the PEP measures the error rate for PSMs with a given score, *x*. Percolator computes the FDR from the PEPs, because the expected number of incorrect PSMs in a given set is equal to the sum of the PEPs[8].

Which Score Is Best?

Suppose that you have just finished running a mass spectrometry experiment, and you have used a database search program to match each spectrum to a peptide. You now need to choose a piece of software to assign statistical scores to each of these PSMs. Assume that you have two choices: one program that computes accurate q-values and one program that computes accurate PEPs. Which program should you choose?

The answer depends on what you plan to do with your results[9]. PEPs and q-values are complementary and useful in different situations. The q-value estimates the rate of misclassification among a set of PSMs. If you want to determine which proteins are expressed in a certain cell type under a certain set of conditions, or if your follow-up analysis involves looking at groups of PSMs, the q-value is an appropriate measure. For example, considering all proteins in a known pathway, evaluating enrichment with respect to Gene Ontology categories, or performing experimental validation on a group of proteins involves analyzing groups of PSMs, so you would use the q-value.

If the goal of your experiment is instead to determine the presence of a specific peptide or protein, use the PEP. For example, suppose that you want to determine whether a certain protein is expressed in a certain cell type under a certain set of conditions. Examine the PEPs of your detected PSMs. Likewise, suppose that you have identified a large set of PSMs using a q-value threshold, and among them you identify a single PSM that is intriguing. Before deciding to dedicate significant resources to investigating a single result, you should examine the PEP associated with that PSM. Although the q-value associated with that PSM might be 0.01, the PEP is always greater than or equal to 0.01. In practice, the PEP values for PSMs near the q = 0.01 threshold are likely to be much larger than 1 percent.

Using Percolator in a Workflow

For instructions on setting up false discovery rates by using Percolator, see "Setting Up False Discovery Rates in the Workflow Editor by Using the Percolator Node" on page 421.

[8] Käll, L.; Storey, J. D.; MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res*, **2008**, *7*: 40–44.

[9] Käll, L.; Storey, J. D.; MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res*, **2008**, *7*: 40–44.

Parameters

Table 135 describes the parameters available for the Percolator node.

Table 135. Percolator node parameters

Parameter	Definition
Maximum Delta Cn	Specifies a Delta Cn threshold and filters out all PSMs with a Delta Cn smaller than this value. The minimum value is 0, and the maximum value is 1. The default is 0.05. Delta Cn is the normalized score difference between the currently selected PSM and the highest-scoring PSM for that spectrum. The default value of 0.05 for the Maximum Delta Cn parameter means that from every spectrum, the PSM with the best score is selected, plus the PSMs that have scores with a normalized difference of no more than 5 percent. The Maximum Delta Cn parameter is a kind of flexible rank filter. PSMs that are excluded from processing by Percolator are classified as low-confidence.
Target FDR (Strict)	Specifies a strict target false discovery rate for highly confident peptide hits. The minimum value is 0, and the maximum value is 1. The default is 0.01.
Target FDR (Relaxed)	Specifies a relaxed target false discovery rate for peptide hits with moderate confidence. The minimum value is 0, and the maximum value is 1. The default is 0.05.
Validation based on	Determines which algorithm to use to calculate the score that the validation is based on: • (Default) q-value: Assigns a statistically meaningful q-value to each PSM. For information on the q-value algorithm, see "q-value" on page 688. • PEP: Assigns to each individual PSM the probability of its being incorrect. For information on Posterior Error Probability (PEP), see "Posterior Error Probability (PEP)" on page 688.

PTM Analysis Nodes

There is only one node, phospho*RS*, in the PTM Analysis section of the Workflow Nodes pane.

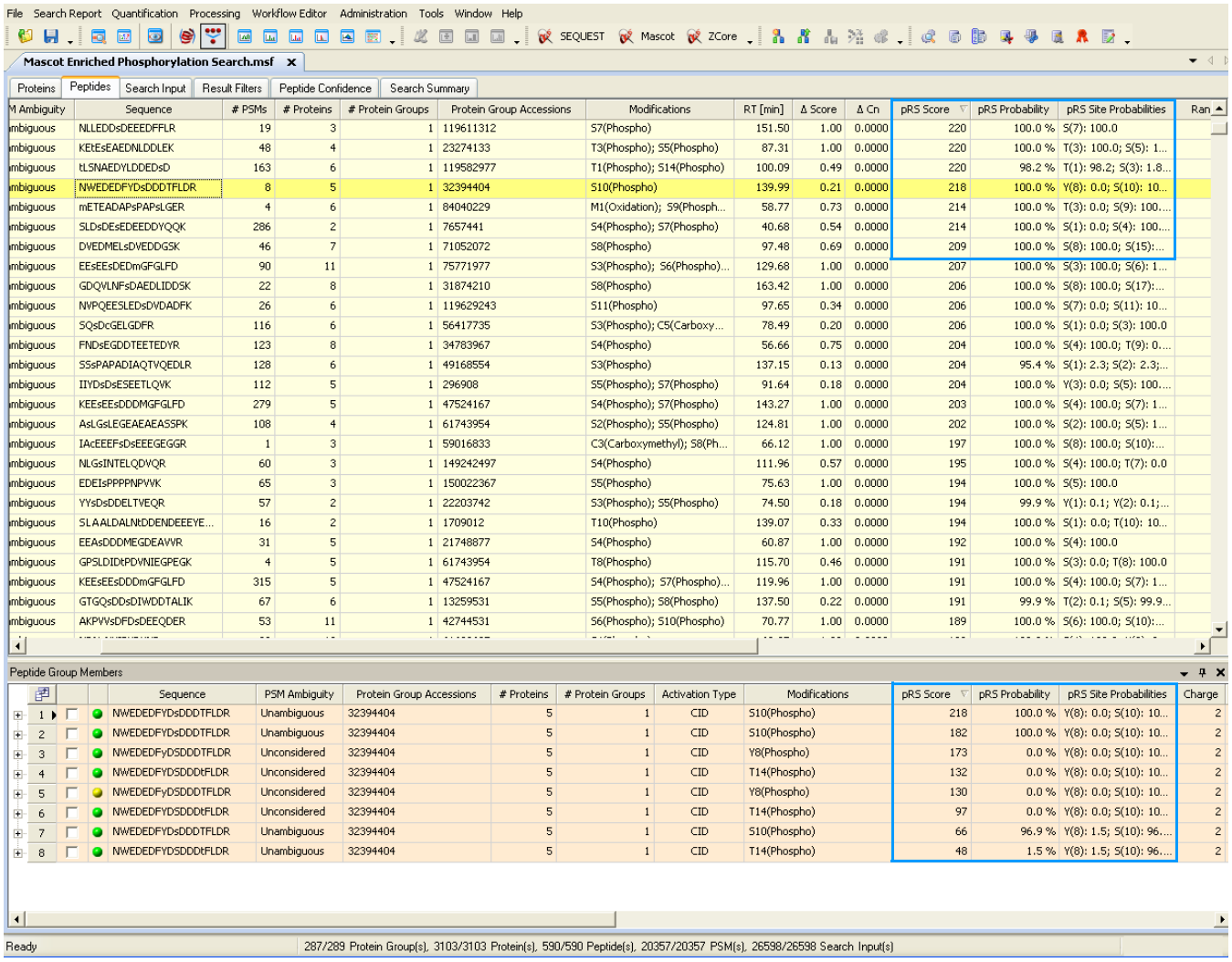
phospho*RS* Node

The phospho*RS* node provides a confidence measure for the localization of phosphorylation in peptide sequences. You can use it with all common fragmentation techniques, such as CID, ETD, and HCD, and peptide identifications from all available database search engines. The input to phospho*RS* is a list of filtered MS/MS fragment spectra with corresponding peptide identifications. For each PSM, phospho*RS* calculates the following:

- pRS score: This peptide score is based on the cumulative binomial probability that the observed match is a random event. The value of the pRS score heavily depends on the data scored, but usually scores above 50 indicate a good PSM.
- pRS sequence probability: This value estimates the probability (0–100%) that the respective isoform is correct.
- pRS site probabilities: For each phosphorylation site, this value is an estimate of the probability 0–100%) that the site is truly phosphorylated. pRS probabilities above 75% give indicate that a site is truly phosphorylated.

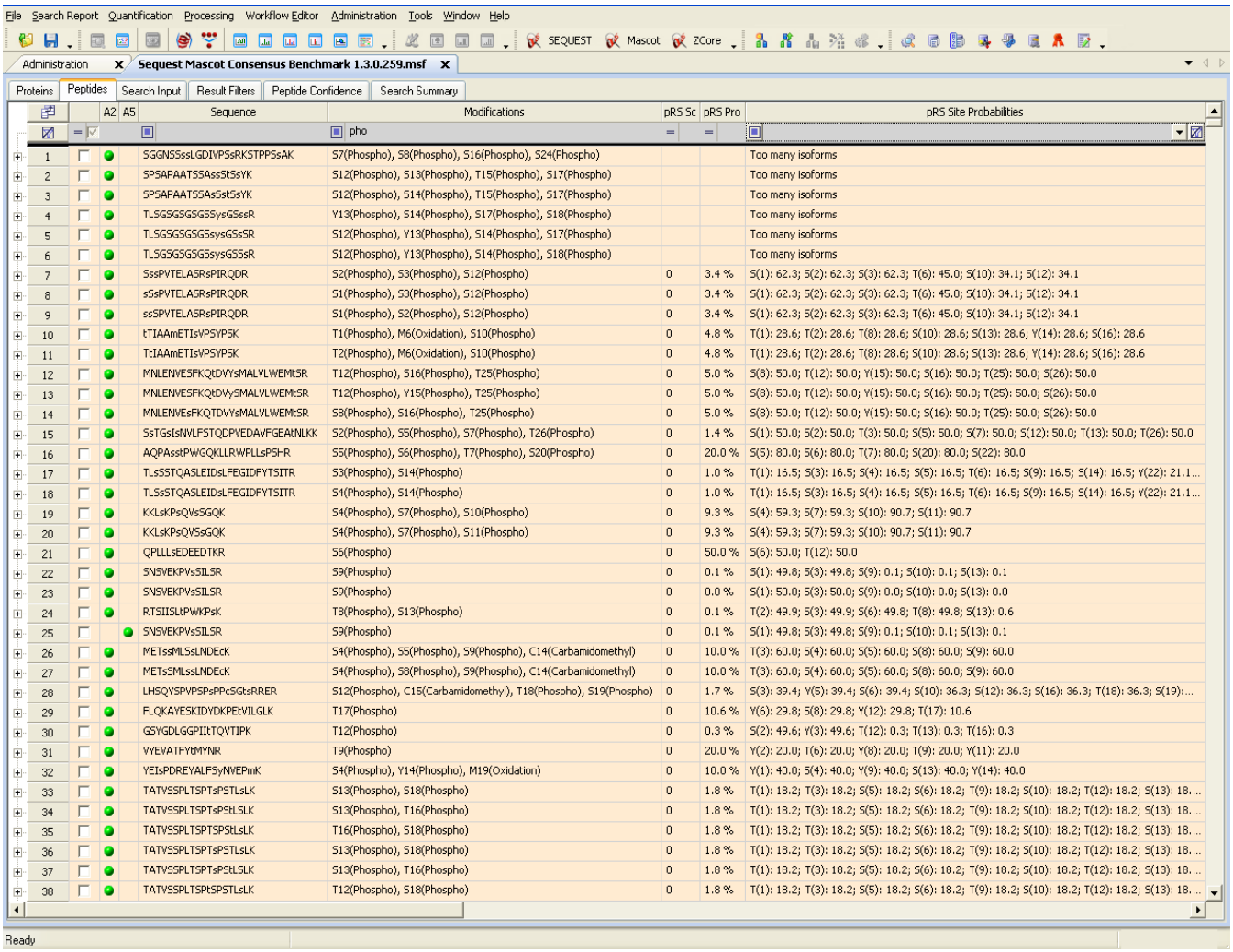
When you open results processed with the phospho*RS* node, the Peptides page shows three values for each phosphorylated peptide: pRS Score, pRS Probability, and pRS Site Probabilities. If you group peptides, the Peptides page shows the scores of the highest-scoring peptide. The Peptide Group Members view displays three additional columns for each PSM. Figure 337 shows these columns.

Figure 337. pRS Score, pRS Probability, and pRS Site Probabilities columns in the Peptides page and the Peptide Group Members view



Note The presence of too many isoforms can prevent phospho*RS* from calculating the pRS site probabilities. In this case, the Proteome Discoverer application displays a message to this effect (see Figure 338) in the pRS Site Probabilities column.

Figure 338. Result of too many isoforms found in a search

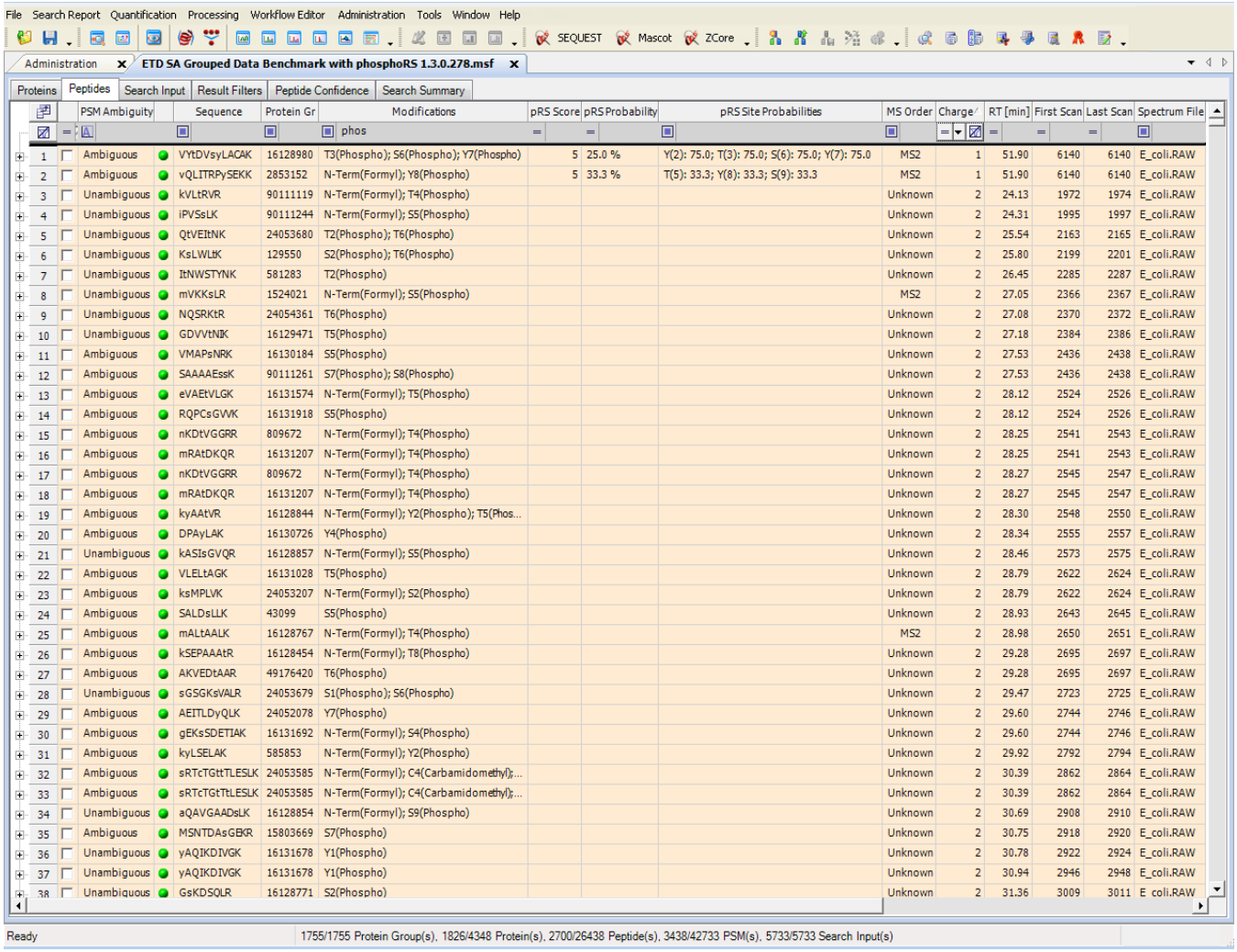


	A2	A5	Sequence	Modifications	pRS Sc	pRS Pro	pRS Site Probabilities
				pho			
1			SGGNSSssLGDIVPSsRKSTPPSsAK	S7(Phospho), S8(Phospho), S16(Phospho), S24(Phospho)			Too many isoforms
2			SPSAPAATSSAssStSsYK	S12(Phospho), S13(Phospho), T15(Phospho), S17(Phospho)			Too many isoforms
3			SPSAPAATSSAsSstSsYK	S12(Phospho), S14(Phospho), T15(Phospho), S17(Phospho)			Too many isoforms
4			TLSGSGSGSGSSysGSssR	Y13(Phospho), S14(Phospho), S17(Phospho), S18(Phospho)			Too many isoforms
5			TLSGSGSGSGSsysGSsSR	S12(Phospho), Y13(Phospho), S14(Phospho), S17(Phospho)			Too many isoforms
6			TLSGSGSGSGSsysGSSsR	S12(Phospho), Y13(Phospho), S14(Phospho), S18(Phospho)			Too many isoforms
7			SssPVTELASRsPIRQDR	S2(Phospho), S3(Phospho), S12(Phospho)	0	3.4 %	S(1): 62.3; S(2): 62.3; S(3): 62.3; T(6): 45.0; S(10): 34.1; S(12): 34.1
8			sSsPVTELASRsPIRQDR	S1(Phospho), S3(Phospho), S12(Phospho)	0	3.4 %	S(1): 62.3; S(2): 62.3; S(3): 62.3; T(6): 45.0; S(10): 34.1; S(12): 34.1
9			ssSPVTELASRsPIRQDR	S1(Phospho), S2(Phospho), S12(Phospho)	0	3.4 %	S(1): 62.3; S(2): 62.3; S(3): 62.3; T(6): 45.0; S(10): 34.1; S(12): 34.1
10			tTIAAmETIsVPSYPSK	T1(Phospho), M6(Oxidation), S10(Phospho)	0	4.8 %	T(1): 28.6; T(2): 28.6; T(8): 28.6; S(10): 28.6; S(13): 28.6; Y(14): 28.6; S(16): 28.6
11			TtIAAmETIsVPSYPSK	T2(Phospho), M6(Oxidation), S10(Phospho)	0	4.8 %	T(1): 28.6; T(2): 28.6; T(8): 28.6; S(10): 28.6; S(13): 28.6; Y(14): 28.6; S(16): 28.6
12			MNLENVESFKQtDVYsMALVLVLWEMtSR	T12(Phospho), S16(Phospho), T25(Phospho)	0	5.0 %	S(8): 50.0; T(12): 50.0; Y(15): 50.0; S(16): 50.0; T(25): 50.0; S(26): 50.0
13			MNLENVESFKQtDVySMALVLVLWEMtSR	T12(Phospho), Y15(Phospho), T25(Phospho)	0	5.0 %	S(8): 50.0; T(12): 50.0; Y(15): 50.0; S(16): 50.0; T(25): 50.0; S(26): 50.0
14			MNLENVEsFKQTDVYsMALVLVLWEMtSR	S8(Phospho), S16(Phospho), T25(Phospho)	0	5.0 %	S(8): 50.0; T(12): 50.0; Y(15): 50.0; S(16): 50.0; T(25): 50.0; S(26): 50.0
15			SsTGsIsNVLFSTQDPVEDAVFGEAtNLKK	S2(Phospho), S5(Phospho), S7(Phospho), T26(Phospho)	0	1.4 %	S(1): 50.0; S(2): 50.0; T(3): 50.0; S(5): 50.0; S(7): 50.0; S(12): 50.0; T(13): 50.0; T(26): 50.0
16			AQPAsstPWGQKLLRWPLLsPSHR	S5(Phospho), S6(Phospho), T7(Phospho), S20(Phospho)	0	20.0 %	S(5): 80.0; S(6): 80.0; T(7): 80.0; S(20): 80.0; S(22): 80.0
17			TLsSTQASLEIDsLFEGIDFYTSITR	S3(Phospho), S14(Phospho)	0	1.0 %	T(1): 16.5; S(3): 16.5; S(4): 16.5; S(5): 16.5; T(6): 16.5; S(9): 16.5; S(14): 16.5; Y(22): 21.1...
18			TLSsSTQASLEIDsLFEGIDFYTSITR	S4(Phospho), S14(Phospho)	0	1.0 %	T(1): 16.5; S(3): 16.5; S(4): 16.5; S(5): 16.5; T(6): 16.5; S(9): 16.5; S(14): 16.5; Y(22): 21.1...
19			KKLsKPsQVsSGQK	S4(Phospho), S7(Phospho), S10(Phospho)	0	9.3 %	S(4): 59.3; S(7): 59.3; S(10): 90.7; S(11): 90.7
20			KKLsKPsQVSsGQK	S4(Phospho), S7(Phospho), S11(Phospho)	0	9.3 %	S(4): 59.3; S(7): 59.3; S(10): 90.7; S(11): 90.7
21			QPLLLsEDEEDTKR	S6(Phospho)	0	50.0 %	S(6): 50.0; T(12): 50.0
22			SNSVEKPVsSILSR	S9(Phospho)	0	0.1 %	S(1): 49.8; S(3): 49.8; S(9): 0.1; S(10): 0.1; S(13): 0.1
23			SNSVEKPVsSILSR	S9(Phospho)	0	0.0 %	S(1): 50.0; S(3): 50.0; S(9): 0.0; S(10): 0.0; S(13): 0.0
24			RTSIISLtPWKPsK	T8(Phospho), S13(Phospho)	0	0.1 %	T(2): 49.9; S(3): 49.9; S(6): 49.8; T(8): 49.8; S(13): 0.6
25			SNSVEKPVsSILSR	S9(Phospho)	0	0.1 %	S(1): 49.8; S(3): 49.8; S(9): 0.1; S(10): 0.1; S(13): 0.1
26			METssMLSsLNDEcK	S4(Phospho), S5(Phospho), S9(Phospho), C14(Carbamidomethyl)	0	10.0 %	T(3): 60.0; S(4): 60.0; S(5): 60.0; S(8): 60.0; S(9): 60.0
27			METsSMLssLNDEcK	S4(Phospho), S8(Phospho), S9(Phospho), C14(Carbamidomethyl)	0	10.0 %	T(3): 60.0; S(4): 60.0; S(5): 60.0; S(8): 60.0; S(9): 60.0
28			LHSQYSPVPSPsPPcSGtsRRER	S12(Phospho), C15(Carbamidomethyl), T18(Phospho), S19(Phospho)	0	1.7 %	S(3): 39.4; Y(5): 39.4; S(6): 39.4; S(10): 36.3; S(12): 36.3; S(16): 36.3; T(18): 36.3; S(19):...
29			FLQKAYESKIDYDKPEtVILGLK	T17(Phospho)	0	10.6 %	Y(6): 29.8; S(8): 29.8; Y(12): 29.8; T(17): 10.6
30			GSYGDLGGPIItTQVTIPK	T12(Phospho)	0	0.3 %	S(2): 49.6; Y(3): 49.6; T(12): 0.3; T(13): 0.3; T(16): 0.3
31			VYEVATFYtMYNR	T9(Phospho)	0	20.0 %	Y(2): 20.0; T(6): 20.0; Y(8): 20.0; T(9): 20.0; Y(11): 20.0
32			YEIsPDREYALFSyNVEPmK	S4(Phospho), Y14(Phospho), M19(Oxidation)	0	10.0 %	Y(1): 40.0; S(4): 40.0; Y(9): 40.0; S(13): 40.0; Y(14): 40.0
33			TATVSSPLTSPTsPSTLsLK	S13(Phospho), S18(Phospho)	0	1.8 %	T(1): 18.2; T(3): 18.2; S(5): 18.2; S(6): 18.2; T(9): 18.2; S(10): 18.2; T(12): 18.2; S(13): 18....
34			TATVSSPLTSPTsPStLSLK	S13(Phospho), T16(Phospho)	0	1.8 %	T(1): 18.2; T(3): 18.2; S(5): 18.2; S(6): 18.2; T(9): 18.2; S(10): 18.2; T(12): 18.2; S(13): 18....
35			TATVSSPLTSPTSPStLsLK	T16(Phospho), S18(Phospho)	0	1.8 %	T(1): 18.2; T(3): 18.2; S(5): 18.2; S(6): 18.2; T(9): 18.2; S(10): 18.2; T(12): 18.2; S(13): 18....
36			TATVSSPLTSPTsPSTLsLK	S13(Phospho), S18(Phospho)	0	1.8 %	T(1): 18.2; T(3): 18.2; S(5): 18.2; S(6): 18.2; T(9): 18.2; S(10): 18.2; T(12): 18.2; S(13): 18....
37			TATVSSPLTSPTsPStLSLK	S13(Phospho), T16(Phospho)	0	1.8 %	T(1): 18.2; T(3): 18.2; S(5): 18.2; S(6): 18.2; T(9): 18.2; S(10): 18.2; T(12): 18.2; S(13): 18....
38			TATVSSPLTSPtSPSTLsLK	T12(Phospho), S18(Phospho)	0	1.8 %	T(1): 18.2; T(3): 18.2; S(5): 18.2; S(6): 18.2; T(9): 18.2; S(10): 18.2; T(12): 18.2; S(13): 18....

In addition, the phospho*RS* node makes PTM site probabilities available in the Protein Identification Details view. For details, see "Interpreting Your Results with the Protein Identification Details View" on page 245.

You cannot use the phospho*RS* node with spectra grouped so that MS orders or activation types or both are mixed. Figure 339 shows the results of this type of grouped data submitted to the phospho*RS* node. In this example, data appears in the pRS Score, pRS Probability, and pRS Site Probabilities columns only for the ungrouped peptides in the first two rows. However, you can use the phospho*RS* node when spectra are grouped so that MS orders and activation types are the same.

Figure 339. Results of grouped data submitted to the phospho*RS* node



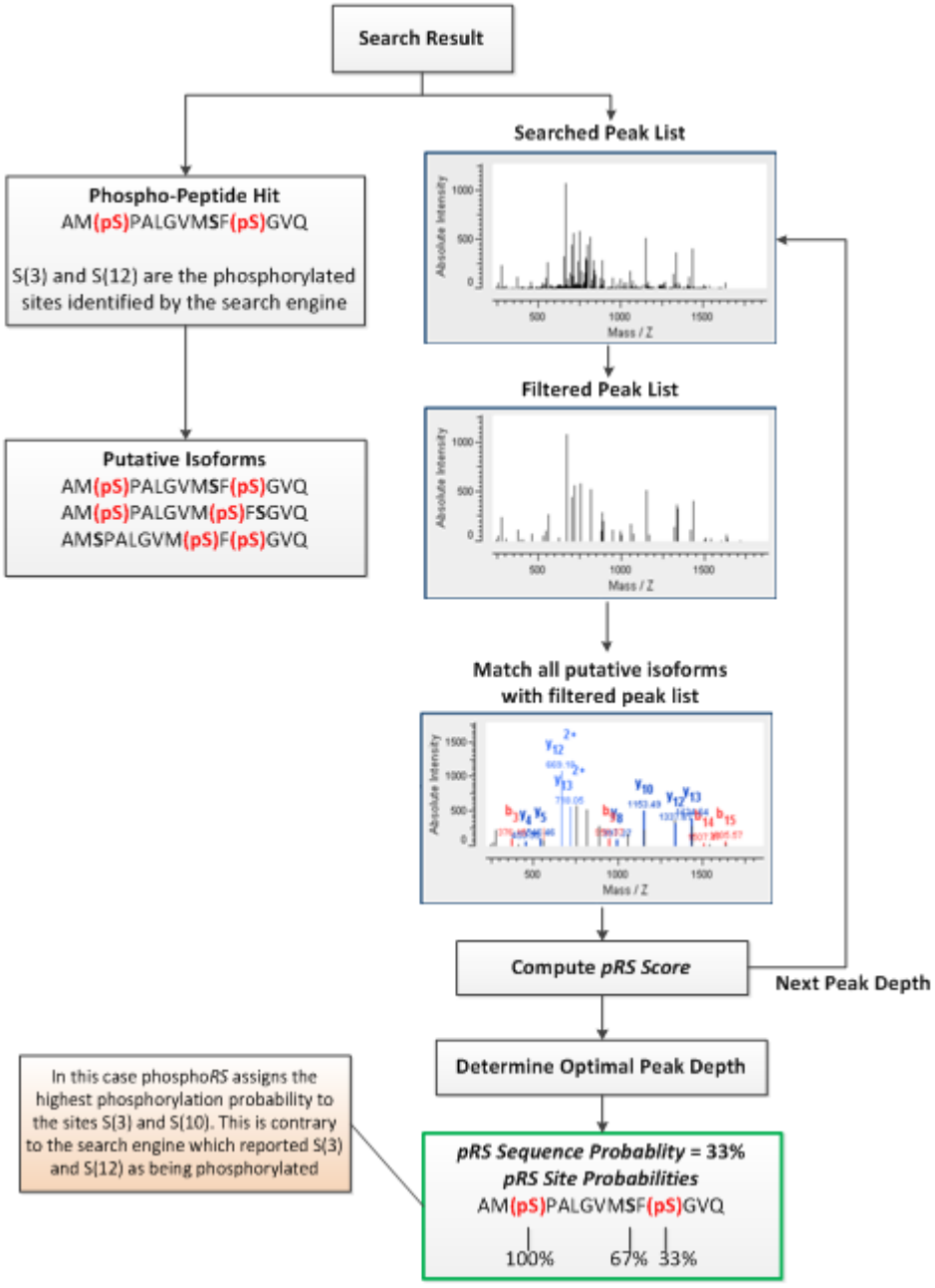
	PSM Ambiguity	Sequence	Protein Gr	Modifications	pRS Score	pRS Probability	pRS Site Probabilities	MS Order	Charge	RT [min]	First Scan	Last Scan	Spectrum File
				phos									
1	Ambiguous	VYtDVsyLACAK	16128980	T3(Phospho); S6(Phospho); Y7(Phospho)	5	25.0 %	Y(2): 75.0; T(3): 75.0; S(6): 75.0; Y(7): 75.0	MS2	1	51.90	6140	6140	E_coli.RAW
2	Ambiguous	vQLITRPySEKK	2853152	N-Term(Formyl); Y8(Phospho)	5	33.3 %	T(5): 33.3; Y(8): 33.3; S(9): 33.3	MS2	1	51.90	6140	6140	E_coli.RAW
3	Unambiguous	kVLtRVR	90111119	N-Term(Formyl); T4(Phospho)				Unknown	2	24.13	1972	1974	E_coli.RAW
4	Unambiguous	iPVSsLK	90111244	N-Term(Formyl); S5(Phospho)				Unknown	2	24.31	1995	1997	E_coli.RAW
5	Unambiguous	QtVEItNK	24053680	T2(Phospho); T6(Phospho)				Unknown	2	25.54	2163	2165	E_coli.RAW
6	Unambiguous	KsLWLtK	129550	S2(Phospho); T6(Phospho)				Unknown	2	25.80	2199	2201	E_coli.RAW
7	Unambiguous	ItNWSTYNK	581283	T2(Phospho)				Unknown	2	26.45	2285	2287	E_coli.RAW
8	Unambiguous	mVKKsLR	1524021	N-Term(Formyl); S5(Phospho)				MS2	2	27.05	2366	2367	E_coli.RAW
9	Unambiguous	NQSRKtR	24054361	T6(Phospho)				Unknown	2	27.08	2370	2372	E_coli.RAW
10	Unambiguous	GDVVtNIK	16129471	T5(Phospho)				Unknown	2	27.18	2384	2386	E_coli.RAW
11	Ambiguous	VMAPsNRK	16130184	S5(Phospho)				Unknown	2	27.53	2436	2438	E_coli.RAW
12	Ambiguous	SAAAAEssK	90111261	S7(Phospho); S8(Phospho)				Unknown	2	27.53	2436	2438	E_coli.RAW
13	Ambiguous	eVAEtVLGK	16131574	N-Term(Formyl); T5(Phospho)				Unknown	2	28.12	2524	2526	E_coli.RAW
14	Ambiguous	RQPCsGVVK	16131918	S5(Phospho)				Unknown	2	28.12	2524	2526	E_coli.RAW
15	Ambiguous	nKDtVGGRR	809672	N-Term(Formyl); T4(Phospho)				Unknown	2	28.25	2541	2543	E_coli.RAW
16	Ambiguous	mRAtDKQR	16131207	N-Term(Formyl); T4(Phospho)				Unknown	2	28.25	2541	2543	E_coli.RAW
17	Ambiguous	nKDtVGGRR	809672	N-Term(Formyl); T4(Phospho)				Unknown	2	28.27	2545	2547	E_coli.RAW
18	Ambiguous	mRAtDKQR	16131207	N-Term(Formyl); T4(Phospho)				Unknown	2	28.27	2545	2547	E_coli.RAW
19	Ambiguous	kyAAtVR	16128844	N-Term(Formyl); Y2(Phospho); T5(Phos...				Unknown	2	28.30	2548	2550	E_coli.RAW
20	Ambiguous	DPAyLAK	16130726	Y4(Phospho)				Unknown	2	28.34	2555	2557	E_coli.RAW
21	Unambiguous	kASIsGVQR	16128857	N-Term(Formyl); S5(Phospho)				Unknown	2	28.46	2573	2575	E_coli.RAW
22	Ambiguous	VLELtAGK	16131028	T5(Phospho)				Unknown	2	28.79	2622	2624	E_coli.RAW
23	Ambiguous	ksMPLVK	24053207	N-Term(Formyl); S2(Phospho)				Unknown	2	28.79	2622	2624	E_coli.RAW
24	Ambiguous	SALDsLLK	43099	S5(Phospho)				Unknown	2	28.93	2643	2645	E_coli.RAW
25	Ambiguous	mALtAALK	16128767	N-Term(Formyl); T4(Phospho)				MS2	2	28.98	2650	2651	E_coli.RAW
26	Ambiguous	kSEPAAAtR	16128454	N-Term(Formyl); T8(Phospho)				Unknown	2	29.28	2695	2697	E_coli.RAW
27	Ambiguous	AKVEDtAAR	49176420	T6(Phospho)				Unknown	2	29.28	2695	2697	E_coli.RAW
28	Unambiguous	sGSGKsVALR	24053679	S1(Phospho); S6(Phospho)				Unknown	2	29.47	2723	2725	E_coli.RAW
29	Ambiguous	AEITLDyQLK	24052078	Y7(Phospho)				Unknown	2	29.60	2744	2746	E_coli.RAW
30	Ambiguous	gEKsSDETIAK	16131692	N-Term(Formyl); S4(Phospho)				Unknown	2	29.60	2744	2746	E_coli.RAW
31	Ambiguous	kyLSELAK	585853	N-Term(Formyl); Y2(Phospho)				Unknown	2	29.92	2792	2794	E_coli.RAW
32	Ambiguous	sRTcTGttTLESLK	24053585	N-Term(Formyl); C4(Carbamidomethyl);...				Unknown	2	30.39	2862	2864	E_coli.RAW
33	Ambiguous	sRTcTGtTtLESLK	24053585	N-Term(Formyl); C4(Carbamidomethyl);...				Unknown	2	30.39	2862	2864	E_coli.RAW
34	Unambiguous	aQAVGAADsLK	16128854	N-Term(Formyl); S9(Phospho)				Unknown	2	30.69	2908	2910	E_coli.RAW
35	Ambiguous	MSNTDAsGEKR	15803669	S7(Phospho)				Unknown	2	30.75	2918	2920	E_coli.RAW
36	Unambiguous	yAQIKDIVGK	16131678	Y1(Phospho)				Unknown	2	30.78	2922	2924	E_coli.RAW
37	Unambiguous	yAQIKDIVGK	16131678	Y1(Phospho)				Unknown	2	30.94	2946	2948	E_coli.RAW
38	Unambiguous	GsKDSQLR	16128771	S2(Phospho)				Unknown	2	31.36	3009	3011	E_coli.RAW

How phospho*RS* Calculates Sequence and Site Probabilities

The calculation of both phospho*RS* sequence probabilities and phospho*RS* site probabilities is based on the assumption that if the sequence of amino acids assigned to the respective MS/MS spectrum is correct and the MS/MS spectrum acquired did not result from fragmentation of two co-eluting peptides with similar *m/z* ratios, one of the putative isoforms must be the true assignment. Therefore, the sum of all sequence probabilities must be equal to 100 percent. If one or both of these assumptions is not correct, the calculated probabilities might not correlate with the true probabilities for correct phospho-site localization.

Figure 340 shows the basic internal workflow of the phospho*RS* node.

Figure 340. Internal phospho*RS* workflow



phospho*RS* performs the following steps:

1. On the basis of all potential phosphorylation sites in the peptide sequence, phospho*RS* calculates all putative phosphorylation isoforms.
2. phospho*RS* divides the fragment spectrum into 100-Th windows and extracts the *i* most intense peaks for each window, where *i* is the peak depth used for the filtered fragment spectrum.

3. For each phosphorylation isoform, *S*, and each peak depth, *i*, phospho*RS* calculates the list of theoretical fragment ions and matches them to the extracted peaks of the experimental spectrum.

   The probability, *p*, of matching a fragment peak purely by chance is calculated as follows:

   $$p = \frac{N_{Peaks} \times d}{w}$$

   where:

   - N_{Peaks} is the number of extracted peaks.
   - *d* is the mass tolerance for matching peaks to the theoretical fragment ions.
   - *w* is the extracted mass range.

4. From the number of theoretical fragment ions, *n*, the number of matched fragment peaks, *kS*, and the probability, *p*, phospho*RS* calculates the probability, $P_{S,i}$ of matching *kS* or more peaks purely by chance as the cumulative binomial probability of matching *kS* or more peaks in *n* attempts[10,11]:

   $$P_{S,i}(X \geq k_s) = \sum_{k = k_S}^{n} \binom{n}{k} p^k \times (1-p)^{(n-k)}$$

5. phospho*RS* reports the pRS score of a specific isoform at peak depth, *i*, as follows:

   $$pRS_{S,i} = 10 \times \log_{10}(P_{S,i}(X \geq k_S))$$

6. phospho*RS* determines the optimal peak depth $i_{optimal}$ as the peak depth with the largest difference between the best-scoring phosphorylation isoform and the second-best-scoring isoform[10].

7. From the pRS score at the optimal peak depth, phospho*RS* calculates a pRS sequence probability and pRS site probabilities. It calculates the pRS sequence probability as follows[11]:

   $$\text{pRS sequence probability} = \frac{P_{S,i_{optimal}}^{-1}}{\sum_S P_{S,i_{optimal}}^{-1}}$$

   Table 136 gives an example showing how phospho*RS* calculates pRS sequence probabilities.

[10]Beausoleil S.A., et al. *Nat. Biotech*. **2006**, *Volume 24*: 1285-1292.
[11]Olsen J.V., et al. *Cell*. **2006**, *127*: 635-648.

Table 136. Calculating pRS sequence probabilities

Putative isoform	pRS score	1/P value	pRS sequence probability
AM(pS)PALGVM(pS)FSGVQ	121	1.26×10^{12}	$1.26 \times 10^{12}/1.89 \times 10^{12}$
AM(pS)PALGVMSF(pS)GVQ	118	6.31×10^{12}	0.33
AMSPALGVM(pS)F(pS)GVQ	59	7.94×10^{12}	0.00
		$\Sigma = 1.89 \times 10^{12}$	$\Sigma = 1.00$

8. phospho*RS* calculates the pRS site probability for a particular phosphorylation site by summing up the pRS sequence probabilities of those isoforms in which the respective site is phosphorylated[12]. Table 137 gives an example showing how phospho*RS* calculates pRS site probabilities.

Table 137. Calculating pRS site probabilities

pRS sequence probability	Putative isoform						
0.67	AM	(pS)	PALGVM	(pS)	F	S	GVQ
0.33	AM	(pS)	PALGVM	S	F	(pS)	GVQ
0.00	AM	S	PALGVM	(pS)	F	(pS)	GVQ
pRS site probabilities		1.00		0.67		0.33	

Using phospho*RS* in a Workflow

For information on creating a workflow that includes the phospho*RS* node, see "Creating a PTM Analysis Workflow" on page 103.

[12]Olsen J.V., et al. *Cell*. **2006**, *127*: 635-648.

phospho*RS* Parameters

Table 138 describes the parameters available for the phospho*RS* node.

Table 138. phospho*RS* node parameters

Parameter	Definition
Fragment Mass Tolerance	Specifies the mass tolerance used for matching fragment ion peaks. **Note** The Proteome Discoverer application uses the set mass tolerance only when you set the Use Fragment Mass Tolerance of Search Node parameter to False. The minimum value is 0.0001 Da, and the maximum value is 1.0 Da. The default is 0.5 Da.
Use Fragment Mass Tolerance of Search Node	Determines whether the Proteome Discoverer application uses the fragment mass tolerance set in the search node for matching fragment ion peaks or the fragment mass tolerance specified by the Fragment Mass Tolerance parameter. • (Default) True: Uses the fragment mass tolerance set in the search node. • False: Uses the fragment mass tolerance specified by the Fragment Mass Tolerance parameter.

Quantification Nodes

The Quantification nodes are available in the Quantification area of the Workflow Editor.

- Precursor Ions Area Detector Node
- Precursor Ions Quantifier Node
- Reporter Ions Quantifier Node

Precursor Ions Area Detector Node

The Precursor Ions Area Detector node calculates the area of each precursor ion. For greater accuracy, it uses an average of the three most abundant peptides rather than the actual peptides to calculate the protein area. You might want to use this node to obtain an idea of the relative quantities of all peptides in a sample.

You cannot use the Precursor Ions Area Detector node and the Precursor Ions Quantifier node in the same workflow.

The Precursor Ions Area Detector node has no parameters.

Precursor Ions Quantifier Node

The Precursor Ions Quantifier node performs quantification for ions that are isotopically labeled. The Proteome Discoverer application uses this node in SILAC experiments.

You cannot use the Precursor Ions Area Detector node and the Precursor Ions Quantifier node in the same workflow.

Table 139 describes the parameters available for the Precursor Ions Quantifier node.

Table 139. Precursor Ions Quantifier node parameters

Parameter	Definition
Quantification Method	Activates the Quantification Method Editor dialog box, shown in Figure 257 on page 493 through Figure 262 on page 501, where you can select a quantification method for precursor ion quantification.
RT Tolerance of Isotope Pattern Multiplets [min]	Specifies the maximum retention-time tolerance of the A0 peaks in the isotope pattern of a quantification multiplet, in minutes. The minimum value is 0.05, and the maximum value is 1. The default value is 0.2 minutes.
Single-Peak/Missing Channels Allowed	Specifies the maximum number of missing quantification channels or quantification channels with just one peak that are allowed for a valid peptide quantification result. The minimum value is 0, and there is no maximum value. The default value is 1.

Reporter Ions Quantifier Node

The Reporter Ions Quantifier node performs reporter ion quantification. This node is used in iTRAQ and TMT experiments.

Table 140 describes the parameters available for the Reporter Ions Quantifier node.

Table 140. Reporter Ions Quantifier node parameters (Sheet 1 of 3)

Parameter	Definition
Activation Type	Specifies the fragmentation method used for each spectrum: • CID (Collision-Induced Dissociation) • MPD (Multi-Photon Dissociation) • ECD (Electron Capture Dissociation) • PQD (Pulsed Q Collision Induced Dissociation) • ETD (Electron Transfer Dissociation) • (Default) HCD (High-Energy Collision Dissociation) • Any Activation Type For a description of these fragmentation types, see "Fragmentation Methods" on page 8.

Table 140. Reporter Ions Quantifier node parameters (Sheet 2 of 3)

Parameter	Definition
Integration Method	Specifies which peak to select if more than one peak is found inside the integration window: • (Default) Most Confident Centroid: Defines which one of multiple peaks to select. This option lays a Gaussian curve around the target peak (the tag mass) with a sigma value equal to the mass accuracy or integration window. Then the Gaussian curve normalizes all peaks in the window, and the peak that is the largest after that is considered to be the most confident peak. This method is also used by the Spectrum Selector node in the Workflow Editor to pick the monoisotopic peak from the survey scan. The only difference is that the Spectrum Selector uses a 3-sigma interval, but Most Confident Centroid uses only a 1-sigma interval. This means the Most Confident Centroid will almost always be the largest peak inside the integration window because of the small inclination of the Gaussian curve in the 1-sigma interval. • Most Intense Centroid: Selects the highest peak. • Centroid With Smallest Delta Mass: Selects the peak with the smallest deviation from the theoretical mass. • Centroid Sum: Sums the intensity of all the peaks in the window.
Integration Window Tolerance	Specifies the mass-to-charge window that enables you to look for the reporter peaks. The minimum value is 0.0001 Da or 0.01 ppm, and the maximum value is 0.6 Da or 5000 ppm. The default is 20 ppm.

Table 140. Reporter Ions Quantifier node parameters (Sheet 3 of 3)

Parameter	Definition
Mass Analyzer	Specifies the type of mass spectrometer used in the acquisition of the spectrum: • Ion Trap (ITMS) • (Default) Fourier Transform (FTMS) • Time of Flight (TOFMS) • Single Quad (SQMS) • Triple Quad (TQMS) • Sector Field (SectorMS)
MS Order	Specifies the level of the mass spectrum to be processed, for example, MS/MS or MS3. You can choose from MS2 through MS10. The default is MS2.
Quantification Method	Activates the Quantification Method Editor dialog box, shown in Figure 257 on page 493 through Figure 262 on page 501, where you can select a quantification method for reporter ion quantification.

Data Export Nodes

There is only one Data Export node, the Spectrum Exporter node, available in the Data Export area of the Workflow nodes pane.

Spectrum Exporter Node

The Spectrum Exporter node converts the spectrum from the available input formats, such as raw, MGF, and SRF formats, to DTA, MGF, MZDATA, or MZML output formats for processing elsewhere.

This feature is not available in the search wizards. When you use the Spectrum Exporter with the Spectrum Selector node, Spectrum Grouper node, or both, you can use the results with additional or customized data processing that is required outside of the Proteome Discoverer application.

Table 141 describes the parameters available for the Spectrum Exporter node.

Table 141. Spectrum Exporter node parameters

Parameter	Definition
Export Format	Specifies the format of the exported spectra. The available formats are the following: • Data Archive: Places the exported spectra into DTA zip files, which are files containing MS^n data for single or grouped scans. • Mascot Generic Format (MGF): Places the exported spectra into MGF files, which are mass spectral files produced during Mascot analysis. They contain a list of precursor ions, their fragments, and the masses of the fragments. • MZDATA: Places the exported spectra into MZDATA files, which are common data format files developed by the Human Proteome Organization (HUPO) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line. • MZML files are a combination of MZDATA and MZXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the Seattle Proteome Center at the Institute for Systems Biology (ISB). The Proteome Discoverer application supports version 1.1.0 of the MZML format. See "Exporting Spectra" on page 114 for instructions on exporting spectra.
File name	Specifies the name of the output file that contains the exported data.

Annotation Nodes

The Annotation nodes are available in the Annotation area of the Workflow Nodes pane.

- Annotation Node
- Re-Annotation Node

Annotation Node

The Annotation node retrieves GO database information from ProteinCenter and installs it in Proteome Discoverer results files. This information is displayed in the Molecular Functions, Cellular Components, and Biological Processes columns in the results report in the MSF file. For information on using this node in a workflow, see "Protein Annotation" on page 433.

The Annotation node has no parameters.

Re-Annotation Node

The Re-Annotation node updates existing annotations or annotates existing MSF files that do not yet include annotations.

The Re-Annotation node must be the only node in a workflow. The node takes an existing MSF file as input, retrieves up-to-date annotations for the proteins contained in the MSF file, and stores them in the same MSF file.

For information on using this node in a workflow, see "Reannotating MSF Files" on page 443.

The Re-Annotation node has only one parameter, MSF File Path, which specifies the path and name of the MSF file to add new annotations to or to update existing annotations in. The browse button (...) opens the Select Analysis File dialog box, which you can use to specify the location of the MSF file.

FASTA Reference

This appendix lists the most important FASTA databases and parsing rules that the Proteome Discoverer application uses to obtain protein sequences, accession numbers, and descriptions.

Contents

- FASTA Databases
- Custom Database Support

FASTA Databases

These are the most important FASTA databases that the Proteome Discoverer application uses.

- NCBI
- MSIPI
- IPI
- UniRef100
- SwissProt and TrEMBL
- MSDB

Follow the links given for each database if you would like to download the database and save it to your local machine. Some databases are more time-consuming to load than others.

NCBI

NCBI is a non-redundant database compiled by the NCBI (National Center for Biotechnology Information) as a protein database for Blast searches. It contains nonidentical sequences from GenBank CDS translations, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), and Protein Research Foundation (PRF).

http://blast.ncbi.nlm.nih.gov/Blast.cgi

ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz

A typical NCBI title line follows:

```
>gi|70561|pir||MYHO myoglobin - horse_i|418678|pir||MYHOZ myoglobin -
common zebra (tentative sequence) [MASS=16950]
```

FASTA ID:

- Accession#:gi70561
- Description:myoglobin - horse_i

MSIPI

MSIPI is a database derived from IPI that contains additional information about cSNPs, N-terminal peptides, and known variants in a format suitable for mass spectrometry search engines. MSIPI is produced by the Max Planck Institute for Biochemistry at Martinsried and the University of Southern Denmark. It is distributed by the European Bioinformatics Institute (EBI).

ftp://ftp.ebi.ac.uk/pub/databases/IPI/msipi/current/

A typical MSIPI title line follows:

```
>MSIPI:IPI00000001.2| Gene_Symbol=STAU1 Isoform Long of
Double-stranded RNA-bin ding protein Staufen homolog 1 lng=577 #
CON[595,R,359,A] #
```

FASTA ID:

- Accession#:IPI00000001.2
- Description:Isoform Long of Double-stranded RNA-bin ding protein Staufen homolog 1 lng=577 # CON[595,R,359,A] #

IPI

The International Protein Index (IPI) is compiled by the European Bioinformatics Institute (EBI) to provide a top-level guide to the main databases that describe the human and mouse proteomes: SwissProt, TrEMBL, NCBI RefSeq, and Ensembl.

http://www.ebi.ac.uk/IPI/

ftp://ftp.ebi.ac.uk/pub/databases/IPI/current/

A typical IPI title line follows:

```
>IPI:IPI00685094.1|SWISS-PROT:2KIJ2|ENSEMBL:ENSBTAP00000028878|REFSE:
NP_001073825;XP_593190 Tax_Id=9913 Gene_Symbol=MGC137286;LOC515210
Uncharacterized protein C1orf156 homolog
```

FASTA ID:

- Accession#:IPI00685094.1
- Description:Uncharacterized protein C1orf156 homolog

UniRef100

UniRef, also known as UniProt NREF, is a set of comprehensive protein databases curated by the Universal Protein Resource consortium. UniRef100 contains only nonidentical sequences, and UniRef90, and UniRef50 are non-redundant at a sequence similarity level of 90 percent and 50 percent, respectively.

http://www.ebi.ac.uk/uniref/

ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/uniref/uniref100/

A typical UniRef100 title line follows:

```
>UniRef100_4U9M9 Cluster: 104 kDa microneme-rhoptry antigen precursor;
n=1; Theileria annulata|Rep: 104 kDa microneme-rhoptry antigen
precursor - Theileria annulata
```

FASTA ID:

- Accession#:4U9M9
- Description:Cluster: 104 kDa microneme-rhoptry antigen precursor; n=1; Theileria annulata|Rep: 104 kDa microneme-rhoptry antigen precursor - Theileria annulata

SwissProt and TrEMBL

The SwissProt database is developed by the SwissProt groups at the Swiss Institute of Bioinformatics (SIB) and the European Bioinformatics Institute (EBI).

TrEMBL is a computer-annotated supplement of SwissProt that contains all the translations of EMBL nucleotide sequence entries not yet integrated into SwissProt.

http://www.expasy.org/sprot/

ftp://ftp.expasy.org/databases/uniprot/knowledgebase/uniprot_sprot.fasta.gz

ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/uniprot_trembl.fasta.gz

A typical SwissProt title line follows:

```
>43495|108_SOLLC Protein 108 precursor - Solanum lycopersicum (Tomato)
(Lycopersicon esculentum)
```

FASTA ID:108_SOLLC

- Accession#:43495
- Description:Protein 108 precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)

MSDB

The Mass Spectrometry Protein Sequence Database (MSDB) is compiled by the Clinical and Biomedical Proteomics group at the University of Leeds, using the PIR, TrEMBL, GenBank, SwissProt, and NRL3D source databases.

http://proteomics.leeds.ac.uk/bioinf/msdb.html

ftp://ftp.ncbi.nih.gov/repository/MSDB/

A typical MSDB title line follows:

```
>CBMS Ubiuinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b -
mouse mitochondrion
```

FASTA ID:

- Accession#:CBMS
- Description:Ubiuinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - mouse mitochondrion

Custom Database Support

The Proteome Discoverer application has three “general” parsing rules to support custom sequence database formats. The generic parsing rules are applied only if no other parsing rule matches the given FASTA title line.

- Custom Parsing Rule A
- Custom Parsing Rule B
- Custom Parsing Rule C

Custom Parsing Rule A

The application uses custom parsing rule A if the FASTA ID, the accession#, and the description are separated by a pipe (|) symbol. A typical FASTA title line, which matches this parsing rule, would look like this one:

```
>tr|18FC3|18FC3_HALWD IS1341-type transposase - Halouadratum walsbyi
(strain DSM 16790).
```

FASTA ID:18FC3_HALWD

- Accession#:18FC3
- Description:IS1341-type transposase - Halouadratum walsbyi (strain DSM 16790).

Custom Parsing Rule B

The application uses custom parsing rule B if the accession# and the description are separated by a space. A typical FASTA title line, which matches this parsing rule, would look like this one:

```
>HP0001 hypothetical protein {Helicobacter pylori 26695}
FASTA ID:
```

- `Accession#:HP0001`
- `Description:hypothetical protein {Helicobacter pylori 26695}`

Custom Parsing Rule C

The application uses custom parsing rule C if the FASTA title line only contains the accession#. A typical FASTA title line, which matches this parsing rule, would look like this one:

```
>143B_HUMAN
FASTA ID:
```

- `Accession#:143B_HUMAN`
- `Description:143B_HUMAN`

Chemistry References

The tables in this appendix list amino acid symbols and mass values, enzyme cleavage properties, and the fragment ions used in the Proteome Discoverer application.

Contents

- Amino Acid Mass Values
- Enzyme Cleavage Properties
- Fragment Ions

Amino Acid Mass Values

The Proteome Discoverer application uses the amino acid symbols and mass values listed in Table 142 and Table 143.

Table 142. Amino acid mass values (Sheet 1 of 2)

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Glycine	G	Gly	57.02147	57.0517	C_2H_3NO
Alanine	A	Ala	71.03712	71.0787	C_3H_5NO
Serine	S	Ser	87.03203	87.078	$C_3H_5NO_2$
Proline	P	Pro	97.05277	97.1168	C_5H_7NO
Valine	V	Val	99.06842	99.1328	C_5H_9NO
Threonine	T	Thr	101.04768	101.1051	$C_4H_7NO_2$
Cysteine	C	Cys	103.00919	103.145	C_3H_5NOS
Isoleucine	I	Ile	113.08407	113.1598	$C_6H_{11}NO$
Leucine	L	Leu	113.08407	113.1598	$C_6H_{11}NO$
Asparagine	N	Asn	114.04293	114.1039	$C_4H_6N_2O_2$
Aspartic Acid	D	Asp	115.02695	115.0885	$C_4H_5NO_3$
Glutamine	Q	Gln	128.05858	128.13091	$C_5H_8N_2O_2$
Lysine	K	Lys	128.09497	128.1745	$C_6H_{12}N_2O$

Table 142. Amino acid mass values (Sheet 2 of 2)

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Glutamic Acid	E	Glu	129.0426	129.1156	$C_5H_7NO_3$
Methionine	M	Met	131.0405	131.1994	C_5H_9NOS
Histidine	H	His	137.05891	137.1414	$C_6H_7N_3O$
Phenylalanine	F	Phe	147.06842	147.1772	C_9H_9NO
Arginine	R	Arg	156.10112	156.188	$C_6H_{12}N_4O$
Tyrosine	Y	Tyr	163.06332	163.17661	$C_9H_9NO_2$
Tryptophan	W	Trp	186.07932	186.2141	$C_{11}H_{10}N_2O$

Table 143. Special amino acids

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Avrg. N/D	B	Bnd	114.53494	114.5962	$C_4H_5NO_3$
Avrg. /E	Z	Ze	128.55059	128.62326	$C_5H_7NO_3$
Unknown acid (X)	X	Xxx	110	110	N/A
Pyrrolysine	O	Pyl	237.14772	237.29874	$C_{12}H_{19}N_3O_2$
Seleno Cysteine	U	Sec	150.95309	150.0369	C_3H_5NOSe

Enzyme Cleavage Properties

Table 144 lists the enzymes and reagents with cleavage properties.

Table 144. Cleavage properties of enzymes and reagents (Sheet 1 of 2)

Enzymes/Reagents	Cleaves after	Cleaves before	Except when
Enzymes for digestion			
AspN		D	
Chymotrypsin	F, W, Y, or L		
Chymotrypsin (FWY)	F, W, or Y		P is after F, W, or Y
Clostripain	R		
Elastase	A, L, I, or V		P is after A, L, I, or V
Elastase/Tryp/Chymo	A, L, I, V, K, R, W, F, or Y		P is after Al, L, I, V, K, R, W, F, or Y

Table 144. Cleavage properties of enzymes and reagents (Sheet 2 of 2)

Enzymes/Reagents	Cleaves after	Cleaves before	Except when
GluC	E or D		
LysC	K		
Proline_Endopept	P		
Staph_protease	E		
Trypsin	K or R		P is after K or R
Trypsin (KRLNH)	K, R, L, N, or H		
Trypsin_K	K		P is after K
Trypsin_R	R		P is after R
Chemicals for degradation			
Cyanogen bromide	M		
Iodobenzoate	W		

Fragment Ions

Fragment ions of peptides are produced by several different fragmentation techniques, such as ECD, ETD, CID, higher-energy C-trap dissociation (HCD), and infrared multi-photon dissociation (IRMPD).

As an example, low-energy CID spectra, which are sequence-specific, are generated by MS/MS and ESI. The fragment ion spectra contain peaks of the fragment ions formed by the cleavage of the peptide bond and are used to determine amino acid sequences. A fragment must have at least one charge for it to be detected.

The fragment ions produced are identified according to where they are fragmented in the peptide. A, b, and c fragment ions have a charge on the N-terminal side, and x, y, and z fragment ions have a charge on the C-terminal side. Fragment ions a*, b*, and y* are ions that have lost ammonia (–17 Da), and fragment ions a^o, b^o, and c^o are ions that have lost water (–18 Da). The subscript next to the letter indicates the number of residues in the fragment ion.[1]

Table 145 summarizes the fragment ions used in the Proteome Discoverer application.

[1] For more information about fragment ions and nomenclature, see Roepstorff, P. and Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrum*. **1984**, *11* (11) 601.

Table 145. Fragment ions

Ions	Description
a	A ion with charge on the N-terminal side
b	B ion with charge on the N-terminal side
c	C ion with charge on the N-terminal side
y	Y ion with charge on the C-terminal side
z	Z ion with charge on the C-terminal side
a*	A ion that has lost ammonia (–17 Da)
b*	B ion that has lost ammonia (–17 Da)
y*	Y ion that has lost ammonia (–17 Da)
a^o	A ion that has lost water (–18 Da)
b^o	B ion that has lost water (–18 Da)
c^o	C ion that has lost water (–18 Da)

Migrating from BioWorks to the Proteome Discoverer Application

This appendix describes how to convert existing MS^n data from Thermo Scientific BioWorks formats into Proteome Discoverer formats.

Contents

- Configuring the SRF Importer
- Installing the Software for the Migration
- Converting SRF, DTA, OUT, and Raw Data
- Possible Causes of Failed Migrations
- Known Issues

Configuring the SRF Importer

To modify the peptide confidence threshold values, the first step in the conversion is to configure the SRF Importer utility.

❖ **To configure the SRF Importer**

1. Choose **Administration > Configuration** or click the **Edit Administration** icon,  .
2. In the Configuration section of the Administration page, click **SRF Importer**.

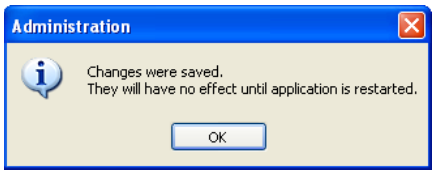
   The SRF Importer view now appears, as shown in Figure 31 on page 66.
3. Modify the peptide confidence threshold values.

   In most cases, the default values are appropriate.
4. If you changed any settings, click 

.

   The message box shown in Figure 341 appears:

Figure 341. Administration message box



5. Click **OK**.

Note Click Reset to return to the default values.

6. Restart your machine.

Installing the Software for the Migration

This section describes how to install the software to move BioWorks SRF data into the Proteome Discoverer application.

Use a computer that can create directories during both BioWorks and the Proteome Discoverer installations without overwriting important files. Table 146 shows the directories that are created during the BioWorks and the Proteome Discoverer installations.

Table 146. Directories created during application installations (Sheet 1 of 2)

Application	Directories created during installation
BioWorks 3.3.1 SP1 (on Windows XP only)	
	C:\Xcalibur\data
	C:\Xcalibur\database
	C:\Xcalibur\examples
	C:\Xcalibur\Install Logs
	C:\Xcalibur\methods
	C:\Xcalibur\params
	C:\Xcalibur\salsa
	C:\Xcalibur\sequence
	C:\Xcalibur\sequest
	C:\Xcalibur\system
	C:\Xcalibur\temp
	C:\Xcalibur\templates
Proteome Discoverer 1.2 (on Windows XP)	

Table 146. Directories created during application installations (Sheet 2 of 2)

Application	Directories created during installation
	C:\Program Files\Thermo\Discoverer
	C:\Program Files\Thermo\Discoverer\ConfigFiles
	C:\Program Files\Thermo\Discoverer\System
	C:\Program Files\Thermo\Discoverer\Tools
Proteome Discoverer 1.2 (on Windows 7)	
	C:\Program Files (x86)\Thermo\Discoverer
	C:\Program Files (x86)\Thermo\Discoverer\ConfigFiles
	C:\Program File (x86)s\Thermo\Discoverer\System
	C:\Program Files (x86)\Thermo\Discoverer\Tools

Preparing the Computer

Install the migration software for data file conversion on a computer that is not used for current data collection, such as a new or unused and empty computer or your desktop computer. Specifically, do not install the data migration software package (BioWorks 3.3.1 SP1 and the Proteome Discoverer conversion workflows) on a computer with the C:\Xcalibur file structure used for instrument control or a computer that already has a licensed BioWorks installation.

Installing BioWorks Software on a Reserved Data System Computer

Note Only Windows XP supports the BioWorks application.

To install the BioWorks 3.3.1 SP1 software for the data migration procedures, you must install the Xcalibur 2.0.7 data system and the BioWorks Browser.

❖ To install the Xcalibur data system and BioWorks Browser

1. Insert the Xcalibur 2.0.7 SP1 CD (or earlier) into your CD-ROM drive.

   The Xcalibur Setup dialog box appears.

2. Click **Xcalibur 2.0.7 Core**.

   The Xcalibur Setup wizard appears.

3. Complete the pages in the installation wizard.

4. Click **Exit**.

5. In your CD-ROM drive, swap the Xcalibur 2.0.7 SP1 CD for the BioWorks 3.3.1 SP1 CD.

   The BioWorks 3.3.1 SP1 setup dialog box appears.

   If your computer already has .NET Framework 2.0, go to step 8.

6. Click **DotNet**.

   The DotNet extraction wizard appears.

7. Complete the pages of the DotNet extraction wizard.

8. Click **BioWorks Browser**.

   The BioWorks Browser Setup wizard appears.

9. Complete the pages of the BioWorks Browser Setup wizard.

   Note On the Database Folder page, choose the specific Xcalibur*database* folder. These FASTA databases require nearly 1 GB of free space to install correctly.

10. Click **Exit**.

Installing the Proteome Discoverer Application

For instructions on installing the Proteome Discoverer application, refer to the *Proteome Discoverer Installation Guide*.

Adding FASTA Databases

For instructions on adding FASTA databases, see "Adding FASTA Files" on page 162.

Installing a Workflow Template

To install the Proteome Discoverer workflow template for the data migration procedures, follow these steps. This template resides in the Example Workflow Templates folder on the Proteome Discoverer installation DVD.

❖ To install the Proteome Discoverer application workflow template

1. Open the application from your desktop by choosing **Start > Programs > Thermo Proteome Discoverer 1.2 > Proteome Discoverer 1.2**.

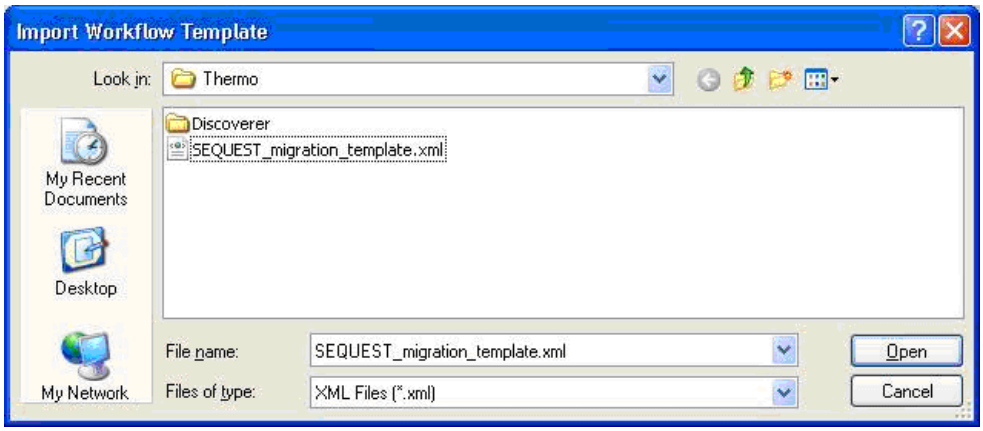
   The application displays an empty startup window.

2. Choose **Workflow Editor > Import Workflow from XML**.

3. Browse to C:\Program Files\Thermo (for Windows XP), as shown in Figure 342, or C:\Program Files (x86)\Thermo\Discoverer (for Windows 7).

4. Select the template XML file, and click **Open**.

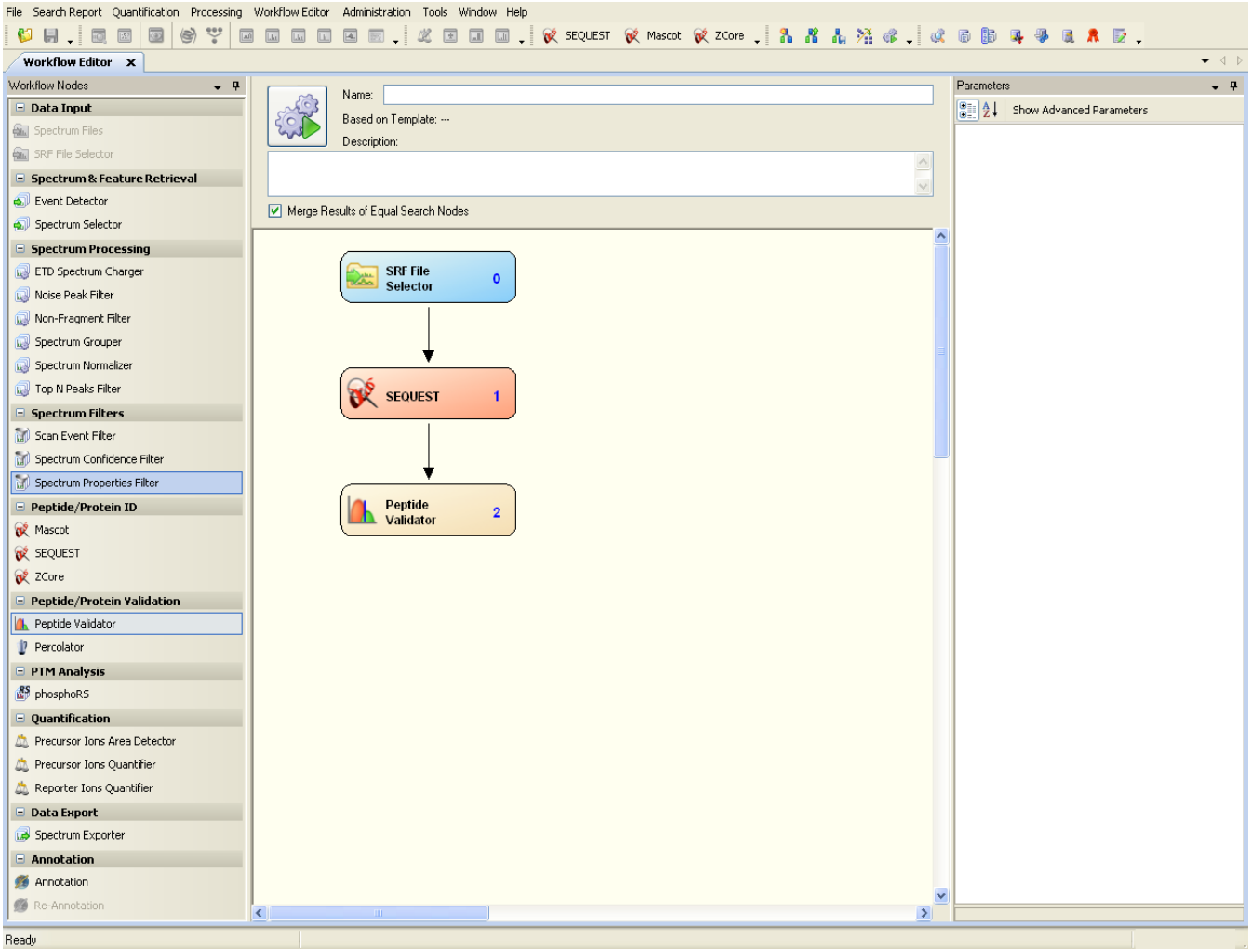
Figure 342. Selecting the template XML file



5. In the Workflow Editor, select the **SRF File Selector** node.

6. In the Parameters pane, click the browse button (...) in the SRF File Path box, as shown in Figure 343, and select the SRF file for migration.

Figure 343. Selecting the SRF file



7. Type a short descriptive name in the Name box.
8. Type a descriptive text string in the Description box.
9. Choose **Workflow Editor > Save As Template**.
10. Click **Save** in the Save Processing Workflow Template dialog box.

Converting SRF, DTA, OUT, and Raw Data

This section describes how to move your MS^n data from the BioWorks to the Proteome Discoverer application. You can use one of two methods to move your BioWorks data:

- Use the SRF File Import Wizard accessible through the File menu, as described in "Moving BioWorks 3.3.1 SRF Files to the Proteome Discoverer Application" on page 721.

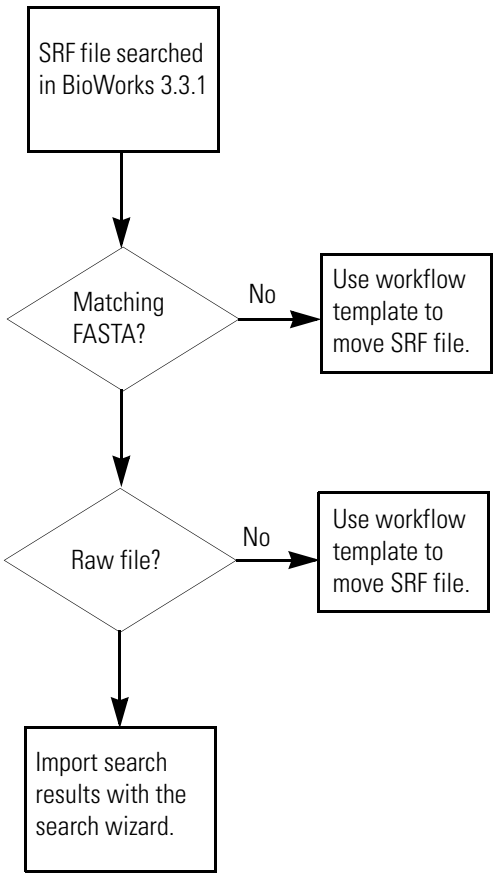
–or–

- Use the SEQUEST migration workflow template that is included in the installation package.

Moving BioWorks 3.3.1 SRF Files to the Proteome Discoverer Application

To import SRF data of BioWorks 3.3.1 origin into the Proteome Discoverer application with the SRF File Import Wizard, you must have access to the original raw and FASTA files. Figure 344 illustrates the procedural flow.

Figure 344. Migration flow diagram for importing BioWorks 3.3.1 SRF files into the Proteome Discoverer application



❖ To import SRF data of BioWorks 3.3.1 origin

1. Open the Proteome Discoverer application.
2. Choose **File > Import Search Results From SRF.**
3. On the Welcome to the SRF File Import Wizard page, click **Next**.

   The File Selection page appears.
4. In the SRF File box, click the browse button (...).

5. Select an SRF file created with BioWorks 3.3.1 from a raw file and a FASTA database file.

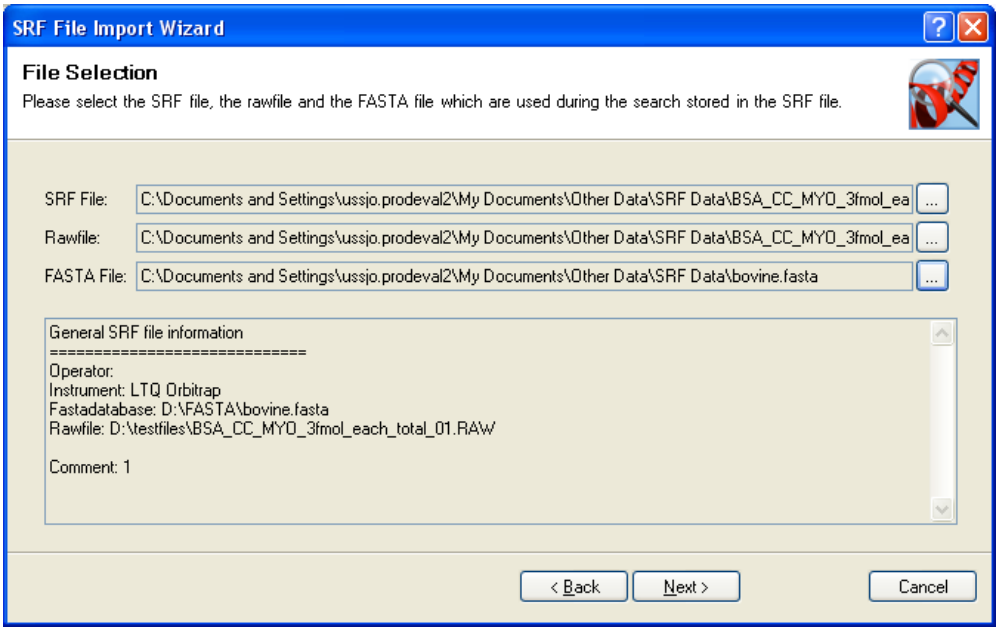
   After you select the SRF file, you can read the information displayed in the header to determine which raw file and which FASTA file were used to create the SRF file.

6. In the Rawfile box, click the browse button (...), and select the raw file listed in the header.

7. In the FASTA File box, click the browse button (...), and select the FASTA file listed in the header.

   The File Selection page of the wizard should now resemble Figure 345.

   Figure 345. File Selection page of the SRF File Import Wizard

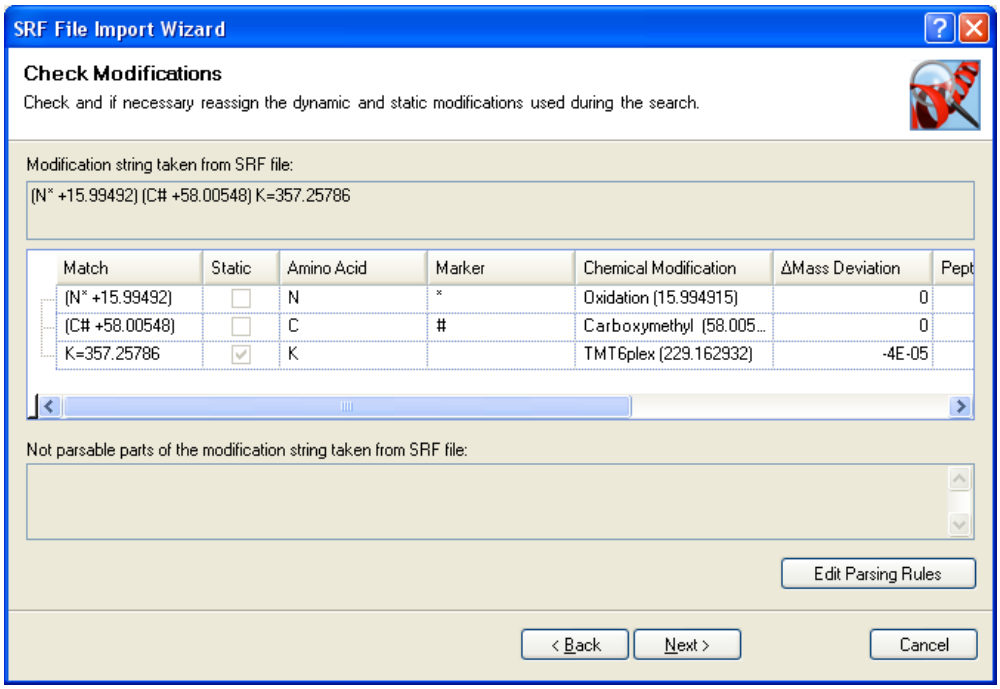


8. Click **Next**.

9. On the Check Modifications page, shown in Figure 346, check the chemical modifications to ensure that they are appropriate.

   A large delta mass, which would be highlighted in red, might indicate that the Proteome Discoverer application selected the wrong chemical modification. For instructions on making other chemical modifications available, see "Chemical Modifications" on page 195.

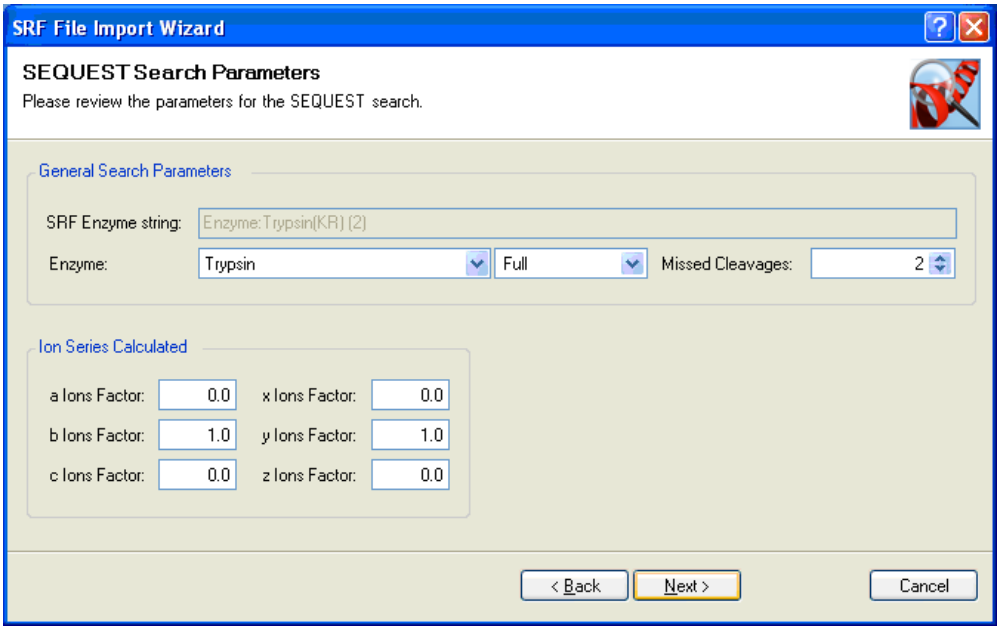
Figure 346. Check Modifications page of the SRF File Import Wizard



10. Click **Next**.
11. On the SEQUEST Search Parameters page, shown in Figure 347, check the search parameters and ion series for the following:
    - The correct enzyme is selected.
    - The number of missed cleavages is correct.
    - The values of the b and y ions are 1 for CID fragmentation.
    - The values of the c, y, and z ions are 1 for ETD fragmentation.

    The information in the Enzyme and Missed Cleavages boxes should match the information in SRF Enzyme string box. If it does not, select the matching information in the Enzyme and Missed Cleavages boxes.

Figure 347. SEQUEST Search Parameters page of the SRF File Import Wizard



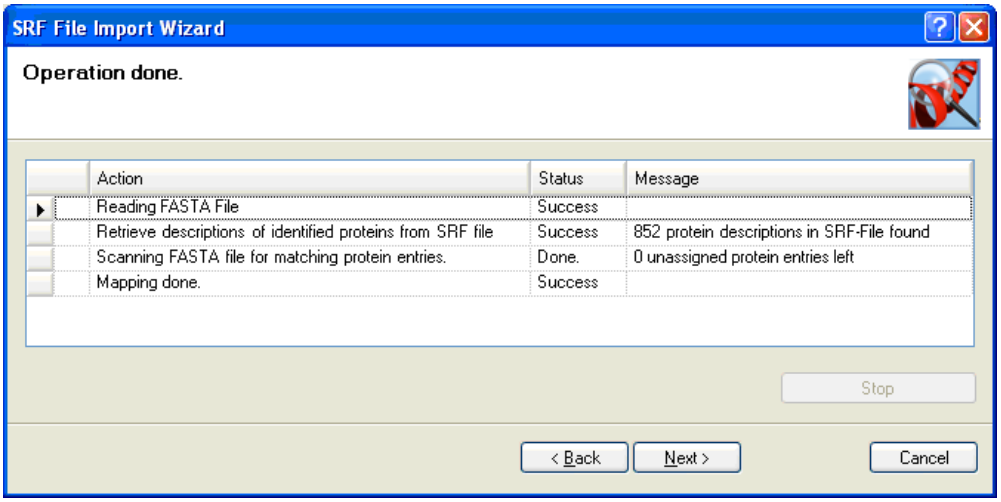
12. Click **Next**.

    The Operation status page appears.

13. After the wizard successfully finishes its operations, click **Next**.

    The Operation Done page appears, as shown in Figure 348.

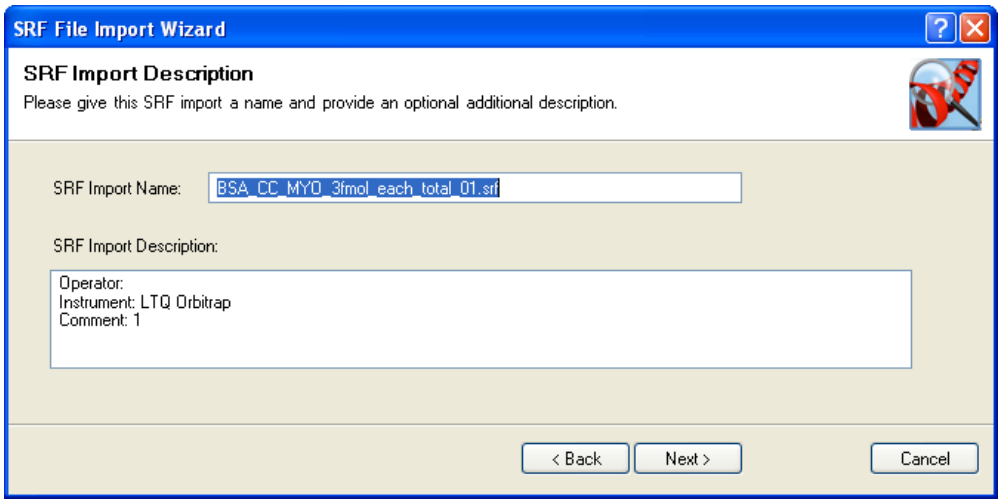
Figure 348. Operation Done page of the SRF File Import Wizard



14. Click **Next**.

    The SRF Import Description page appears, as shown in Figure 349.

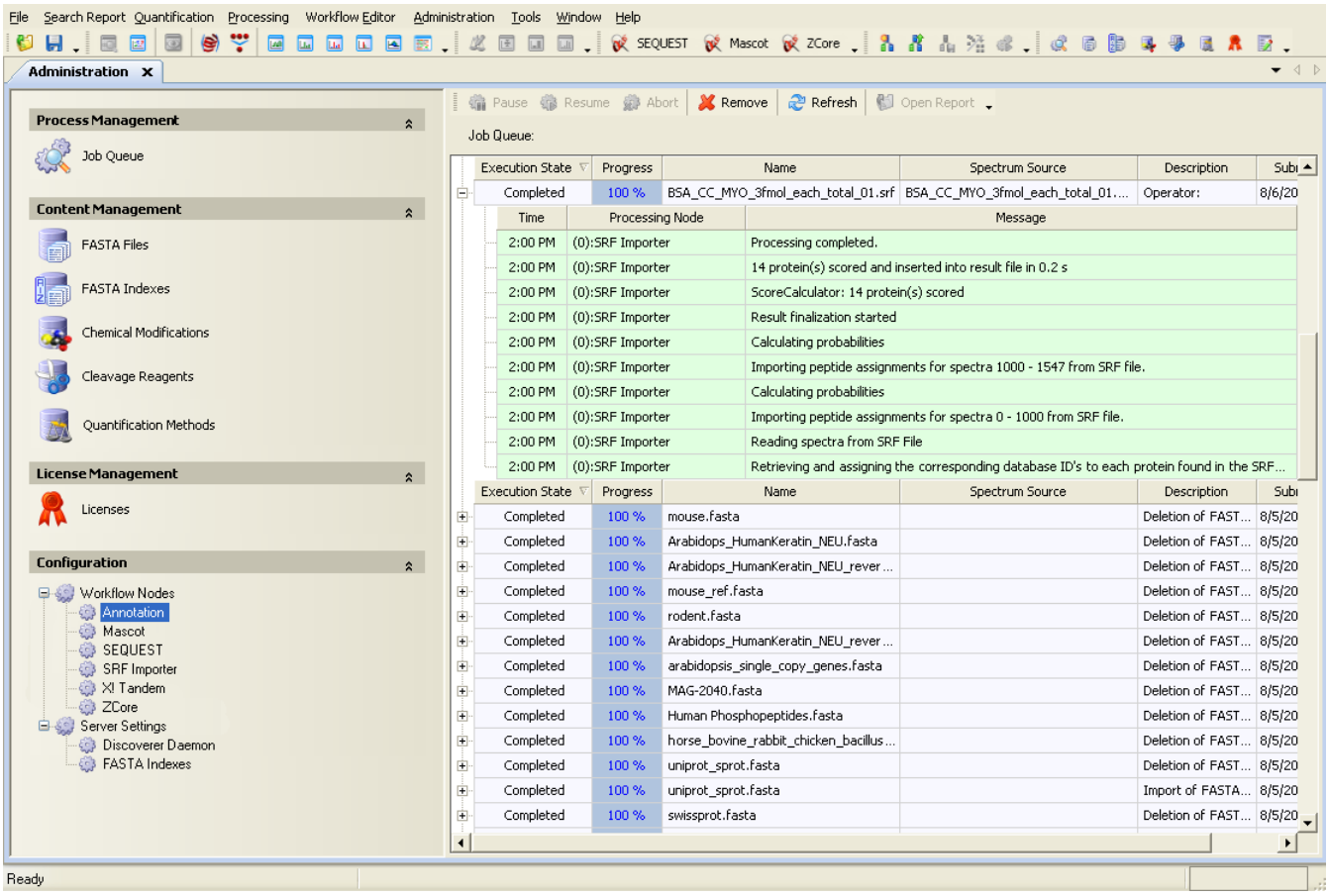
Figure 349. SRF Import Description page of the SRF File Import Wizard



15. In the SRF Import Description box, type a description of the searched file.
16. Click **Next**.
17. (Optional) On the Complete the Wizard page of the wizard, read the summary of the search.
18. Click **OK**.

    The SRF file is imported into the Proteome Discoverer application. The job queue appears, as shown in Figure 350.

Figure 350. Job queue



19. Click the job row to activate the completed job.
20. In the job queue toolbar, click the **Open Report** icon,  .

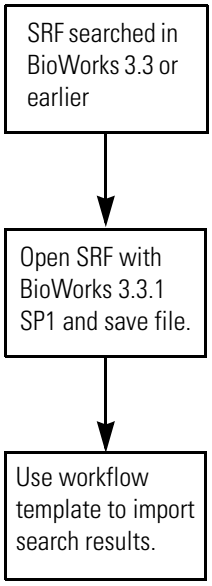
    The Input Files page appears. For information about loading a results report (MSF file), see "Opening MSF Files" on page 215.

For information about the results page, see "Interpreting Search Results" on page 213. If the migration failed, see "Possible Causes of Failed Migrations" on page 737.

Moving BioWorks 3.3 or Earlier SRF Files to the Proteome Discoverer Application

To import SRF data of BioWorks 3.3 origin or of earlier BioWorks releases into the Proteome Discoverer application, use the following procedure while referencing the original FASTA files. Figure 351 illustrates the steps in this procedure.

Figure 351. Migration flow diagram for BioWorks 3.3 or earlier SRF files



❖ To import BioWorks 3.3 or earlier SRF files into the Proteome Discoverer application

Note In this example, information about the raw file, such as which FASTA file was used, comes from the SRF file. This information is necessary to set up the workflow nodes correctly.

1. Open BioWorks 3.3.1 SP1 from your desktop by choosing **Start > Programs > Xcalibur > BioWorks Browser**.

   BioWorks displays an empty startup window.

2. Choose **File > Open Search Results File** (SRF).

   The Open Search Results File dialog box opens.

3. Use the browser functionality of the dialog box to locate the SRF file, and click **Open**.

4. Choose **File > Save Search Results**.

   A dialog box prompts you to either **Save** the converted file under the same name or to **Save as** another name.

5. Save the converted file.

6. Open the Proteome Discoverer application from your desktop by choosing **Start > Programs > Thermo Proteome Discoverer > Proteome Discoverer 1.2**.

7. Choose **Administration > Maintain FASTA Files**.

   If the FASTA files listed in the FASTA files view include the FASTA file that you searched for, go to step 9.

8. In the FASTA files view, click the **Add** icon, 

.

9. Choose the FASTA file used to create the SRF file that originated from the OUT and DTA files.

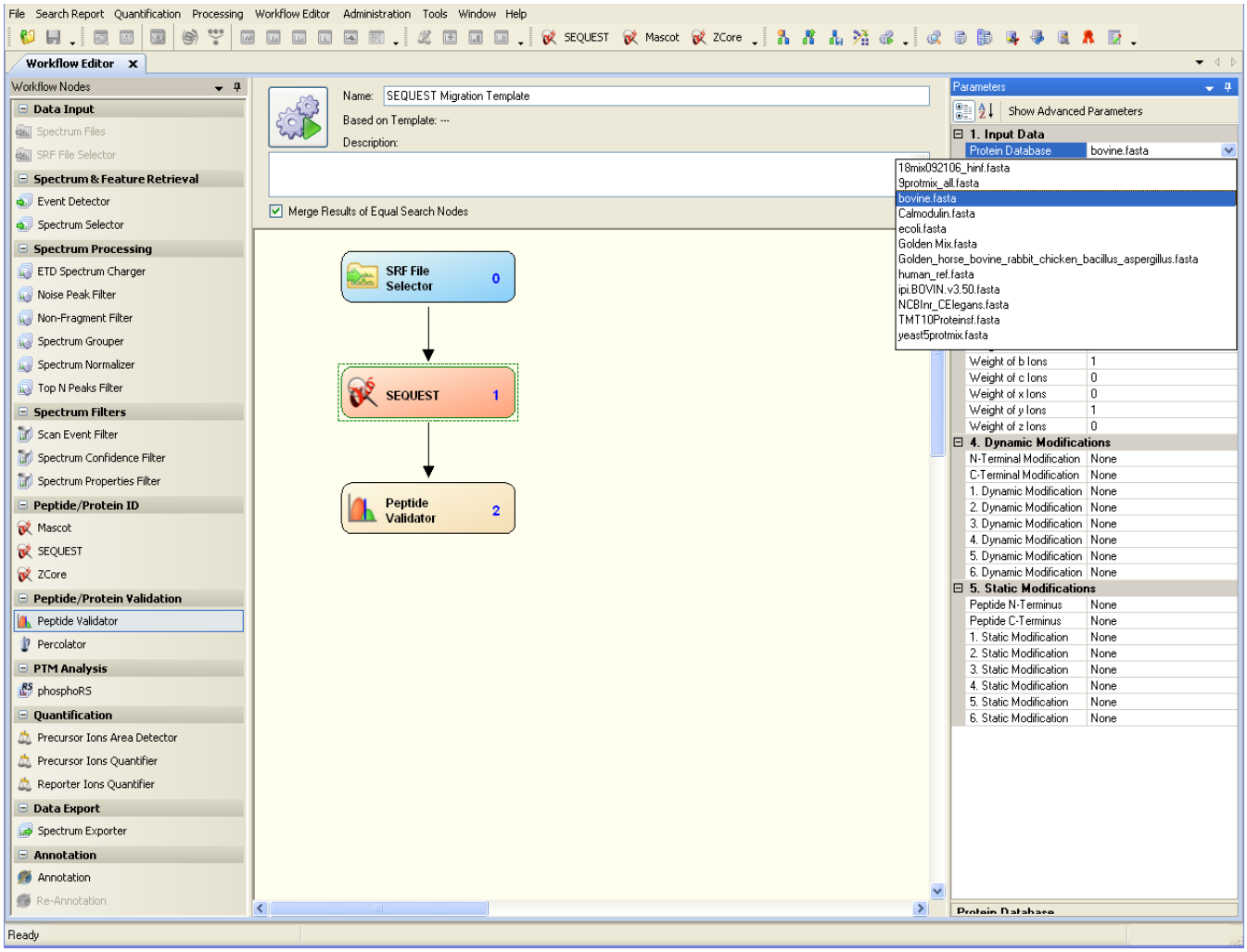
   Note You can open the SRF file in the BioWorks 3.3.1 SP1 version included in this migration package to find out how it was searched.

   –or–

   Make your own migration template using the Workflow Editor and your preferred search engine.

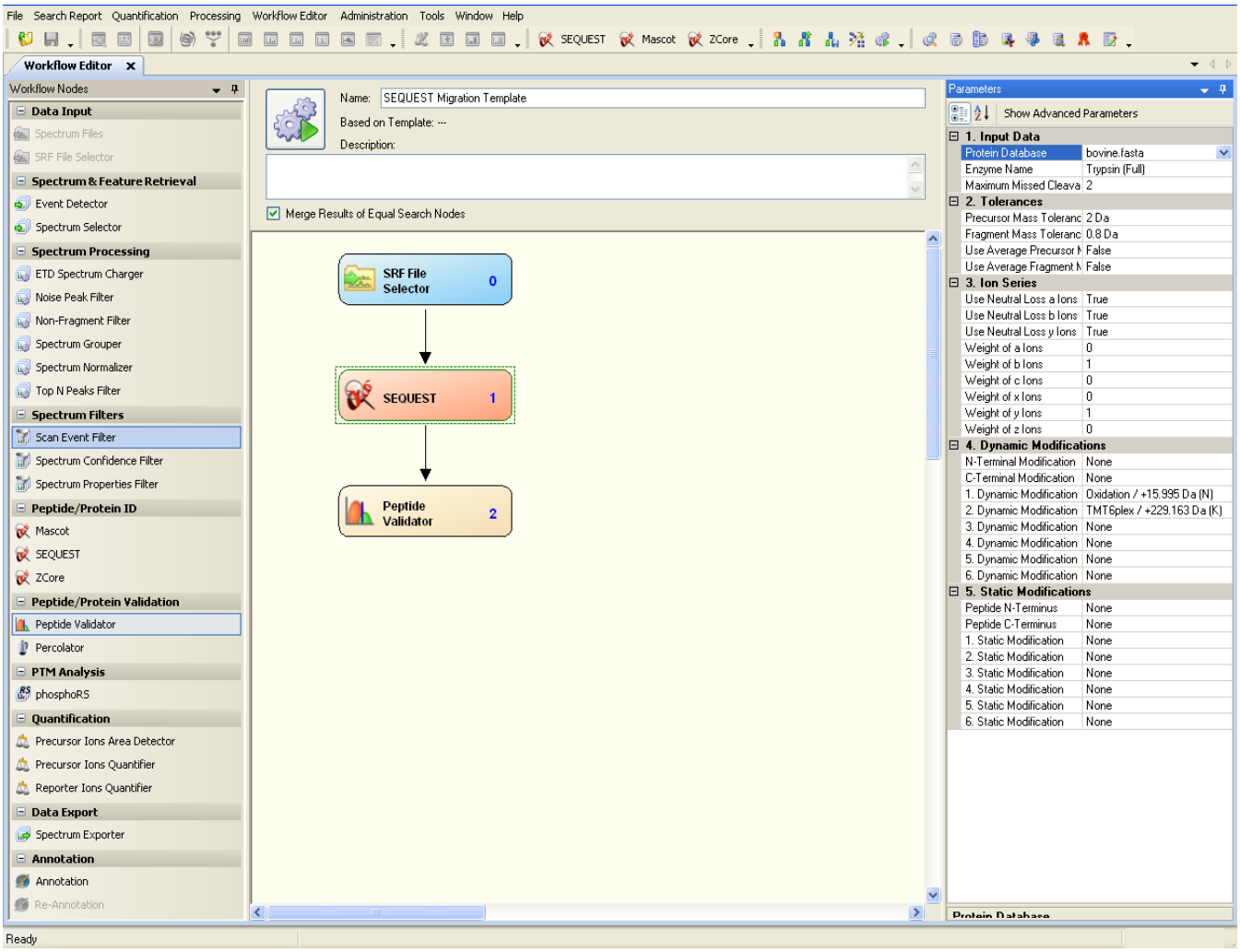
10. Choose **Workflow Editor > Open From Template**.
11. Select the migration template.
12. Click the **SRF File Selector** node.
13. In the Parameters pane, click the browse button (...) in the SRF File Path box, and select the SRF file, as shown in Figure 343 on page 720.
14. Click the **SEQUEST** node.
15. In the Protein Database box in the Parameters pane, click the browse button (...) and select the FASTA file, as shown in Figure 352.

Figure 352. Selecting the FASTA file in the Workflow Editor



16. Change the SEQUEST node parameters to match your data set, as shown in Figure 353.

Figure 353. Changing the SEQUEST node parameters in the Workflow Editor



17. Type a short descriptive name in the Name box.
18. Type a descriptive text string in the Description box.
19. (Optional) Choose **Workflow Editor > Save As Template**.
20. Choose **Workflow Editor > Start Workflow**.

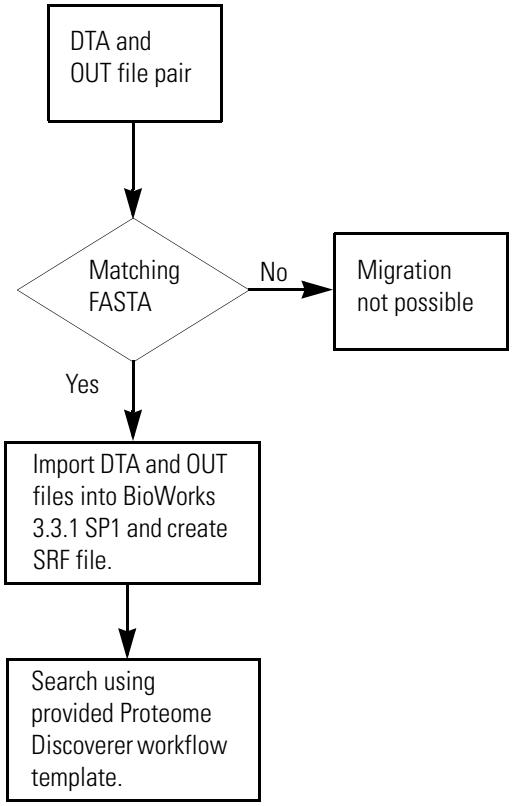
    The job queue appears, showing the status of your search.

Importing BioWorks DTA, OUT, and FASTA Files into the Proteome Discoverer Application

To import a BioWorks DTA and an OUT file while referencing the original FASTA files, use the following procedure. Figure 354 illustrates the procedural flow.

Note You cannot search DTA and OUT files from merged or grouped protein results in the Proteome Discoverer application.

Figure 354. Migration flow diagram for BioWorks DTA, OUT, and FASTA files

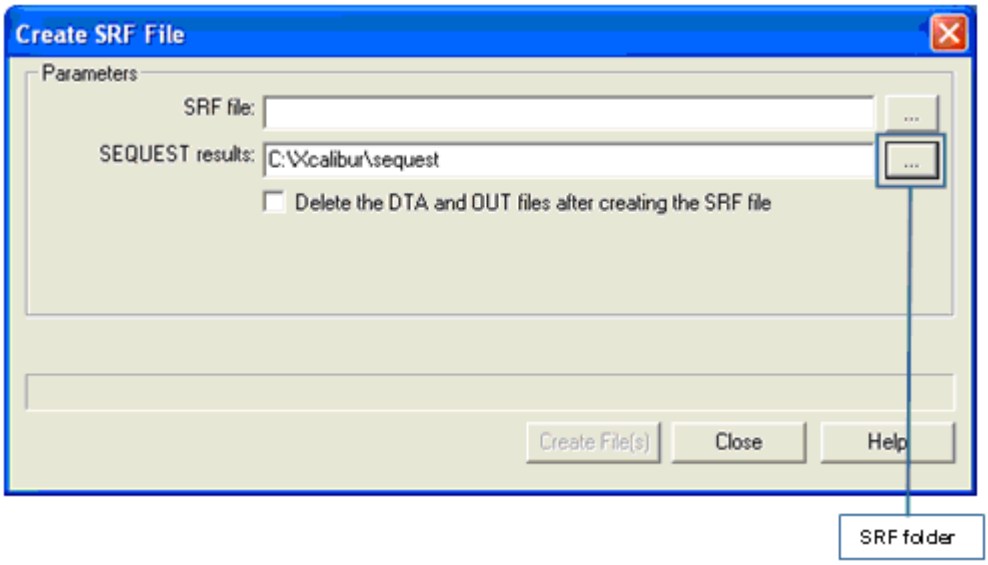


❖ To move BioWorks 3.3 or earlier DTA and OUT files to BioWorks 3.3.1 SP1

1. Open BioWorks 3.3.1 SP1 from your desktop by choosing **Start > Programs > Thermo Xcalibur > BioWorks Browser**.
2. Choose **Tools > Convert > SEQUEST Results > Create SRF File**.

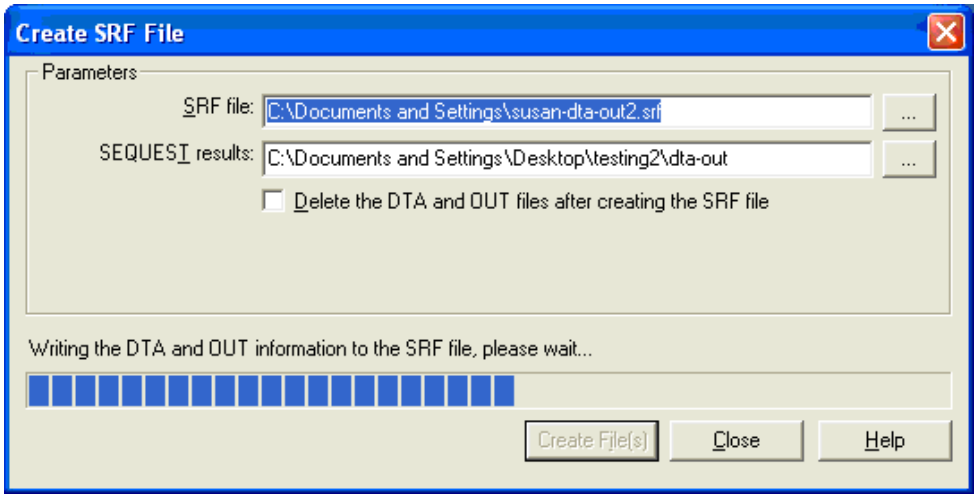
   The Create SRF File dialog box appears, as shown in Figure 355.

Figure 355. Create SRF File dialog box



3. In the SRF file box, type the name and path of the file to save the SRF data in, or click the browse button (...) to browse to the name of an existing file.
4. In the SEQUEST results box, click the browse button (...), and browse to the DTA and OUT file folder, as shown in Figure 356.

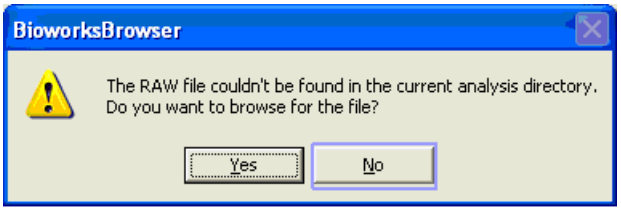
Figure 356. Selecting the DTA and OUT files



5. Click **Create File(s)**.

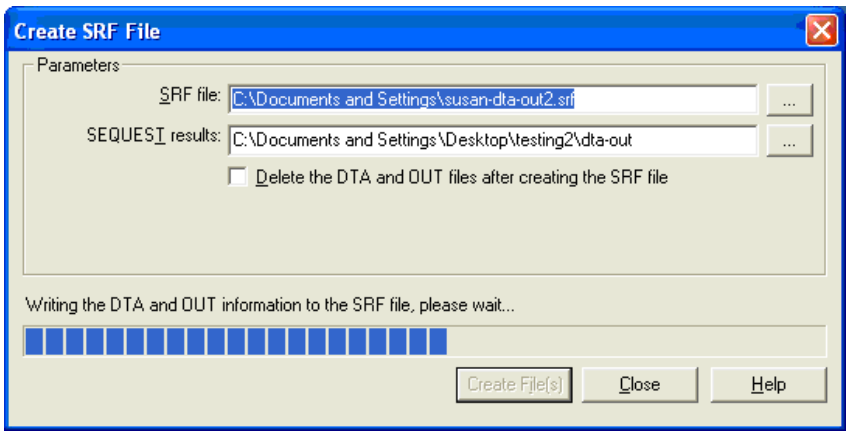
   If the raw file is not in the folder, the BioWorksBrowser prompt box appears, as shown in Figure 357.

Figure 357. BioworksBrowser prompt box



6. Click **No**.
7. After the process begins, deactivate the text box by clicking away from the SRF File text box.
8. Wait several minutes for BioWorks to assemble the SRF file. See Figure 358.

Figure 358. Converting view



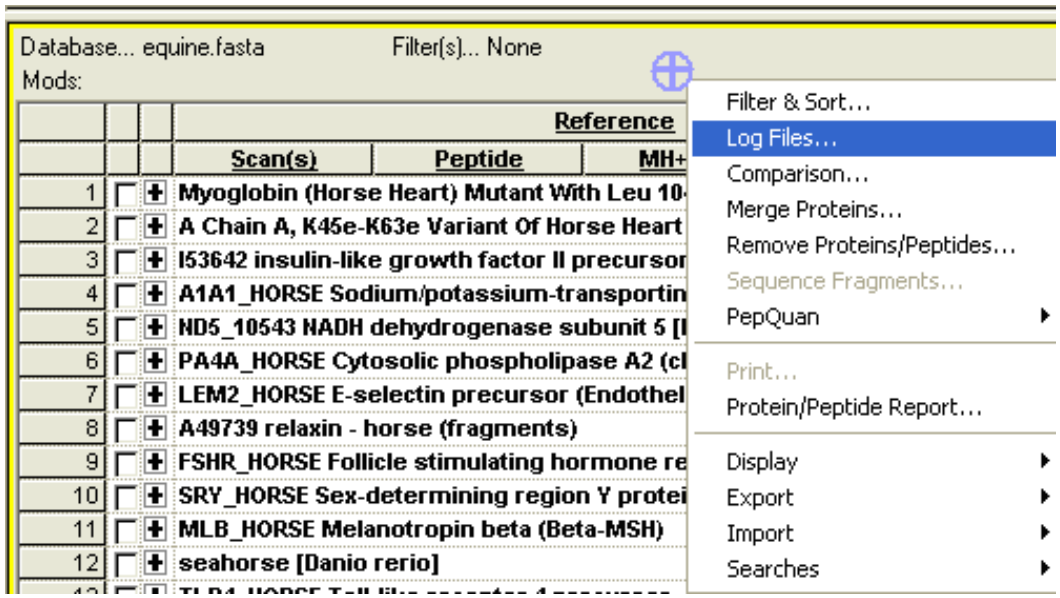
9. Click **Close**.

You now have an SRF file for migration in the Proteome Discoverer application, using an imported Proteome Discoverer workflow template. For details on the installed template, see "Installing a Workflow Template" on page 718.

❖ To verify that the converted SRF file is complete and contains usable data

1. Choose **Open > Open Search Results File (SRF)**.
2. In the protein grid, right-click and choose **Display Peptide Information** from the shortcut menu.
3. In the SRF view, right-click and choose **Log Files**, as shown in Figure 359.

Figure 359. SRF log file information



4. Examine the file information. Make note of the referenced FASTA file.
5. If the information that you viewed during steps 1 and 2 is complete, open the Proteome Discoverer application.

   A correct file has understandable values and contains no null fields or immense numbers in columns of reasonable data. For examples of good data sets, see "Possible Causes of Failed Migrations" on page 737.
6. Update your FASTA files to match the FASTA searched SRF files that you are importing. Choose **Administration > Maintain FASTA files**.
7. In the Administration view menu, click the **Add** icon, 

.
8. Choose the FASTA file used to create the SRF file originating from the OUT and DTA files.
9. Choose **Workflow Editor > Open From Template**, and select the imported SEQUEST migration template for your data set.
10. Click the **SRF File Selector** node, click the browse button (...) in the Parameters pane, and select the SRF file, as shown in Figure 343 on page 720.
11. Click the **SEQUEST** node and examine its parameter settings in the Parameters pane, as shown in Figure 353 on page 730.
12. In the Protein Database box, click the browse button (...) and select the FASTA file for the SEQUEST node.
13. Change the SEQUEST node parameters to fit your data set, as shown in Figure 353 on page 730.
14. Type a short descriptive name in the Name box.

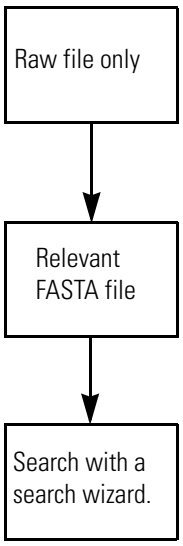
15. Type a descriptive text string in the Description box.
16. (Optional) Choose **Workflow Editor > Save As Template**.
17. Choose **Workflow Editor > Start Workflow**.

    The job queue appears, showing the status of your search.

Importing Raw Data Files into the Proteome Discoverer Application

The following procedure describes how to use a search wizard (in this example, SEQUEST) to import raw data files into the Proteome Discoverer application. Figure 360 shows the procedural flow.

Figure 360. Migration flow diagram for raw data files



❖ To use the SEQUEST search wizard on a raw file

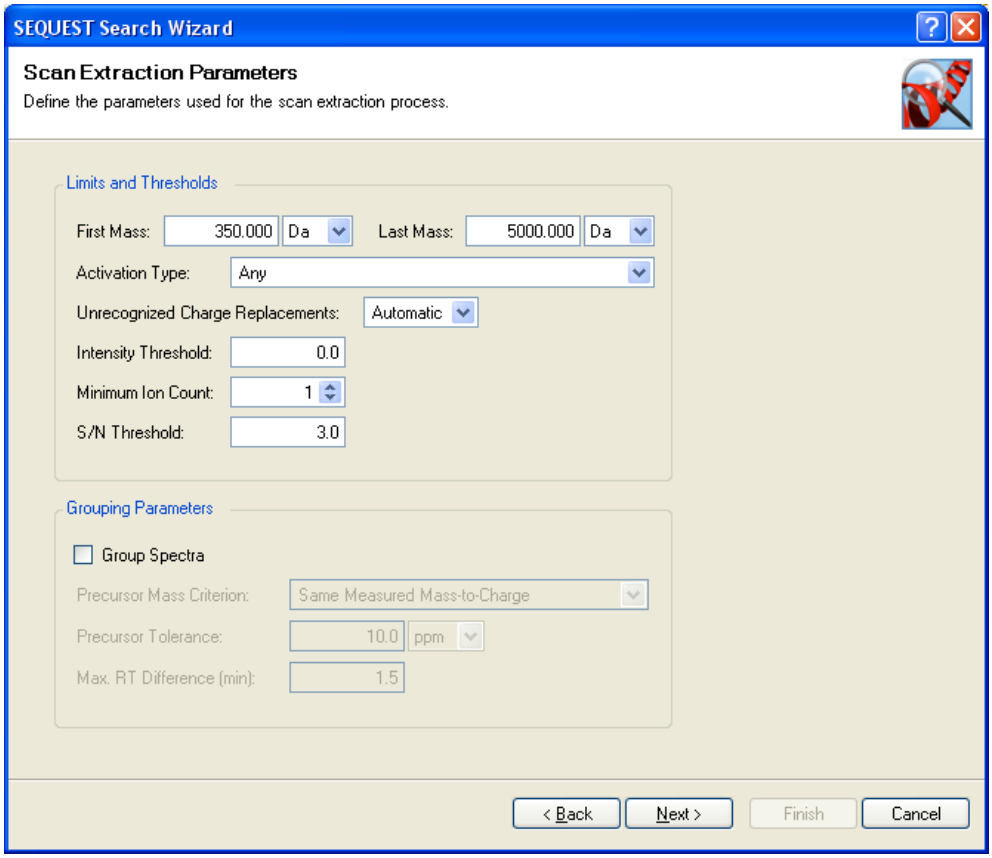
1. Choose **Start > Programs > Thermo Proteome Discoverer 1.2 > Proteome Discoverer 1.2**.
2. Choose **Processing > Start SEQUEST Search Wizard**.

   The SEQUEST Search Wizard dialog box appears.
3. Click **Next**.

   The Rawfile and Scan Range Selection page appears.
4. Click the browse button (...), and select a raw file.
5. To select a scan range, press CTRL and drag the cursor over the range.
6. Click **Next**.

   The Scan Extraction Parameters page appears, as shown in Figure 361.

Figure 361. Scan Extraction Parameters page of the SEQUEST Search Wizard



7. In the Limits and Thresholds area, set the scan extraction parameters:
   a. Enter the first and last mass limits.
   b. Set the mass units.
   c. Select an activation type.
   d. Type a value for the intensity threshold.
   e. Type a value for the minimum ion count.
   f. Type a value for the signal-to-noise threshold setting.
8. (Optional) Determine your grouping parameters.
9. Click **Next**.

   The SEQUEST Search page appears.
10. Set the SEQUEST search parameters:
    a. Select a FASTA database.
    b. Define your enzyme type.
    c. Define the number of missed cleavages.

11. (Optional) Select the Decoy Database Search option and FDR parameters.
12. Set the precursor mass search tolerance parameters:
    a. (Optional) Select the **Use Average Precursor Mass** option.
    b. Define the precursor mass tolerance value.
    c. Select the precursor mass units (**Da**, **mmu**, **ppm**).
13. Set the values in the Ions Series area as necessary for your experiment type.
14. Click **Next**.

    The Modifications page appears.
15. (Optional) Enter the modifications and click **Next**.

    The Search Description page appears.
16. Enter a search name and search description.
17. Click **Next**.

    The completed search results appear.
18. Click **Finish**.

❖ To view your search results

1. Choose **Administration > Show Job Queue**.
2. After the job finishes, click the completed row.
3. On the job queue toolbar, click the **Open Report** icon, 

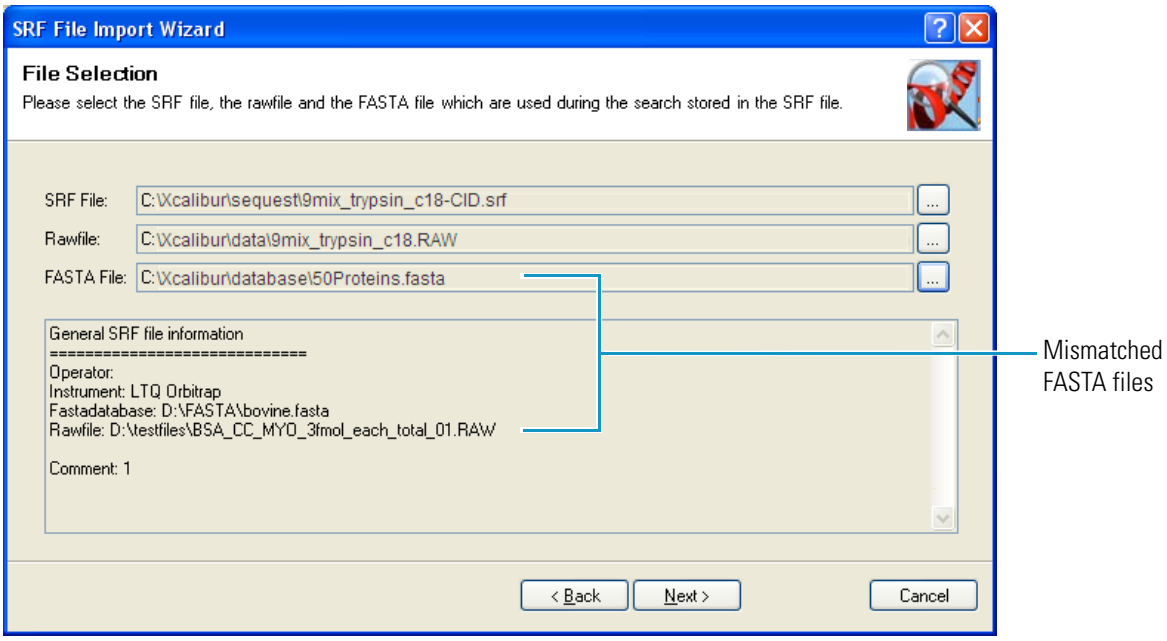
, to display your search results.
4. Use different views to aid in your analysis, such as clicking a protein or peptide row to navigate to other tabular and graphical views.
5. For more help in reviewing your results file, press F1.

Possible Causes of Failed Migrations

A migration can fail for the following reasons:

- Data was saved with Group Scan or CombIon: Group scans require that the DTA files be combined in chromatographic proximity. CombIon combines DTA files with matching spectra by adding them together.
- Merged protein results: Merging protein results might create a mismatch between the identified peptides and the FASTA sequence.
- Wrong FASTA file association, as shown in Figure 362.

Figure 362. Mismatched SRF and FASTA files



- Corrupted SRF, DTA, or OUT files, as shown in Figure 363.

Figure 363. Corrupted BioWorks SRF file

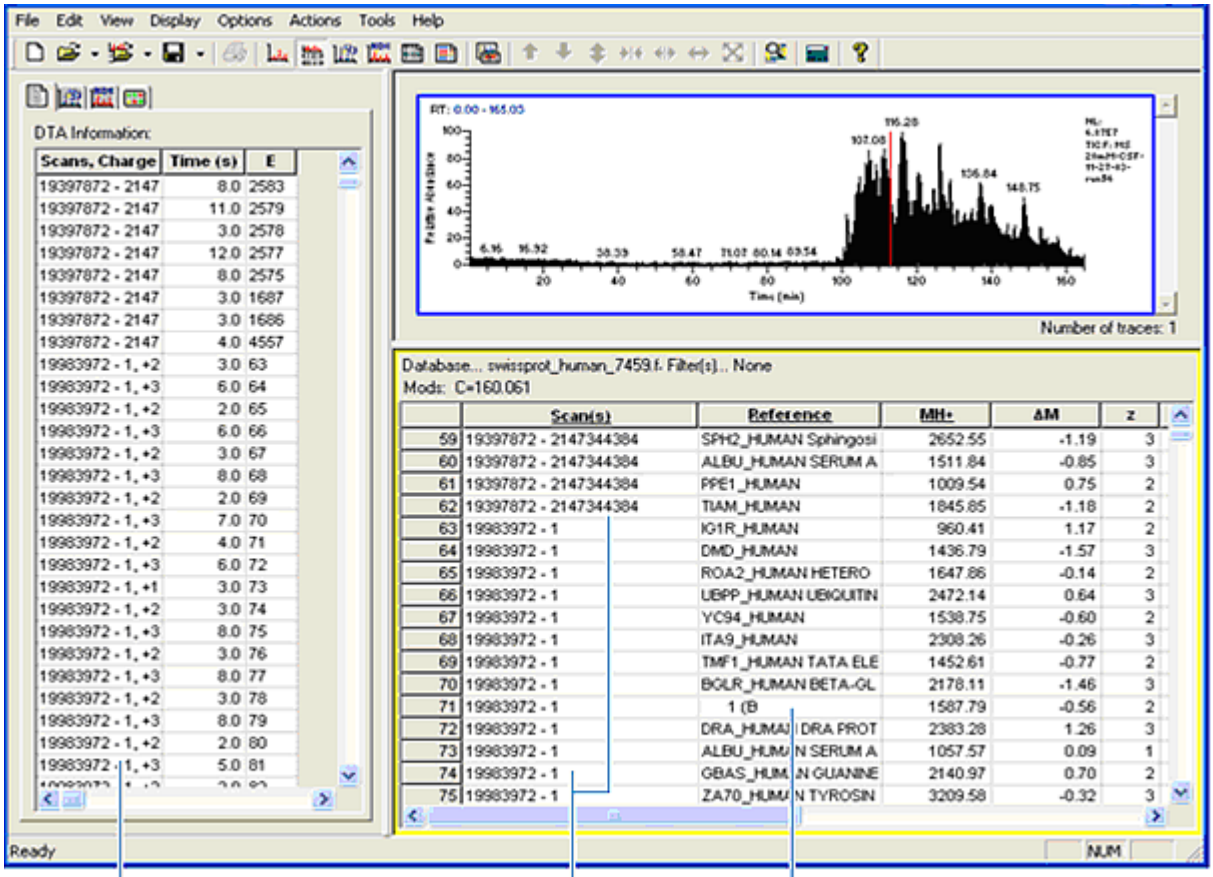
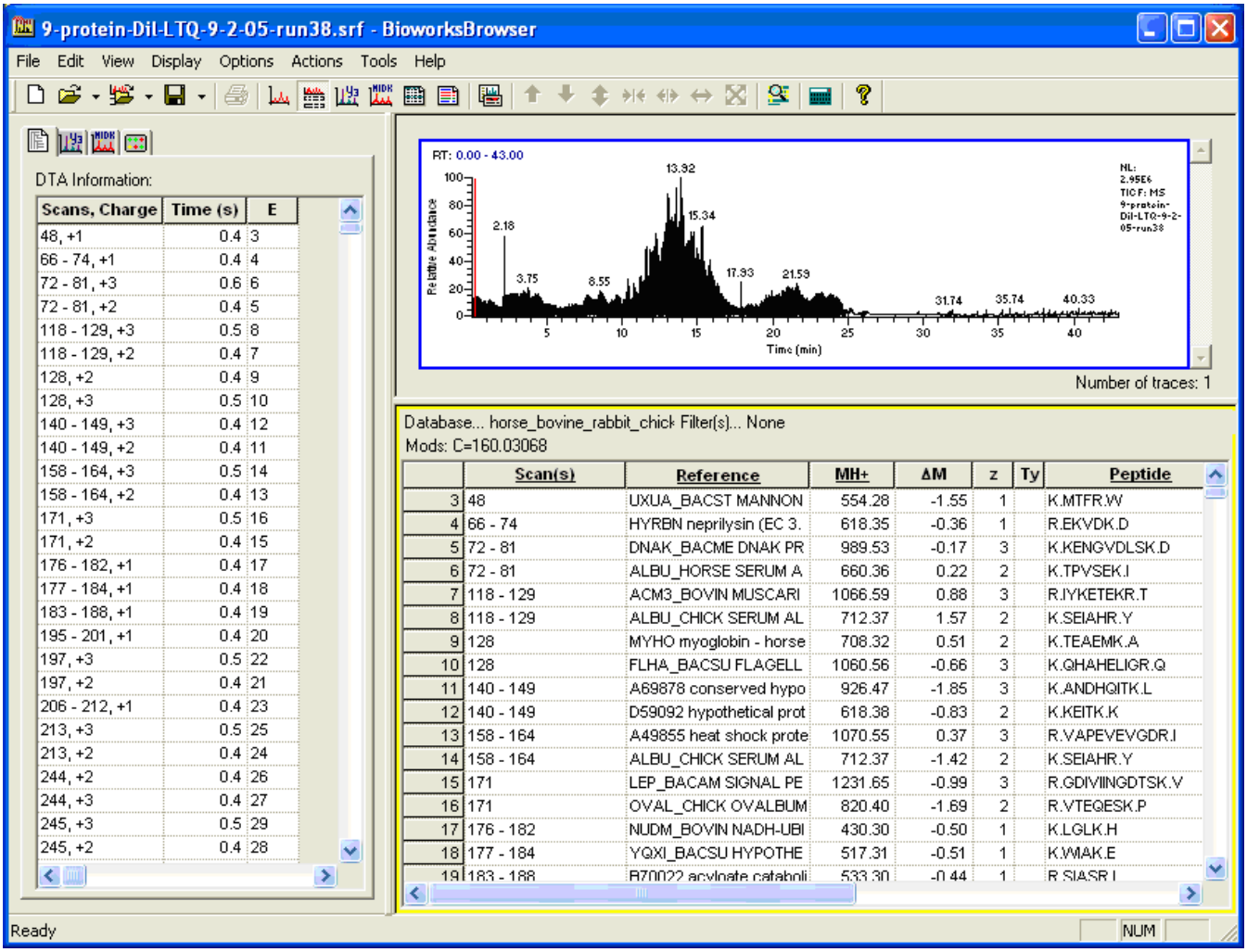


Figure 364 shows a compatible BioWorks SRF file.

Figure 364. Compatible BioWorks SRF file



Known Issues

You might encounter the following known issues when you import SRF files into the Proteome Discoverer application:

- There is a known problem with accessing a raw file within the Proteome Discoverer application that is already opened in the Xcalibur 2.0.7 data system or earlier version. When you try to run a search with the same raw file or display an extracted ion chromatogram in the application, the search fails or the application shows an error. The Proteome Discoverer application already uses raw file access to libraries from the Xcalibur 2.1 data system that is incompatible with raw file access to libraries from previous versions

of the Xcalibur data system. You can run the Proteome Discoverer application with the Xcalibur 2.0.7 data system or earlier, but you must ensure that all raw files opened in the Xcalibur data system are closed before opening them in the Proteome Discoverer application. Thermo Fisher Scientific recommends that you upgrade your current Xcalibur data system installation to version 2.1.

- The ETD Spectrum Charger node is not used for ETD spectra searched with the SEQUEST or Mascot search engine when using the wizards. However, you can simply use the Workflow Editor for this purpose.
- Sometimes when the Proteome Discoverer application crashes unexpectedly, the underlying service program (Thermo.Magellan.Server.exe) does not shut down correctly. In some cases, the application cannot reconnect or restart through the service program. In this case, use the Windows Task Manager to manually end the process. To open the Windows Task Manager, right-click the task bar and then choose Task Manager from the shortcut menu.
- When the search type of the peptides-generated SRF file from the DTA and OUT files is not set (unknown), the search result is empty. Under this condition, all peptides are excluded from the SRF import. The search type is used to distinguish ZCore search results from SEQUEST results. The Proteome Discoverer application requires this difference because it cannot import ZCore search results.

Index

A

a* fragment ions 713–714
a° fragment ions 713–714
a fragment ions
  definition 7, 302, 714
  displayed in Peptide Identification Details view 302, 311, 314
  flag representing 318
  location of charge 300
  used for spectrum matching 676
Abbreviation column 199
Abort icon 51
About Thermo Proteome Discoverer command 629
Accession column 241
accession keys 452
Activation Type column
  Peptides page 293
  Search Input page 347
Activation Type peptide filter
  Mascot reports 371
  SEQUEST reports 366
  ZCore reports 375
activation types 8
Add Analysis File(s) dialog box 219, 352
Add icon
  importing SRF files 727
  opening MSF files 216, 351
  verifying converted SRF files 734
Add Protein References page 165, 175
adding and removing amino acids 201
adding chemical modifications 200
Administration menu 22, 626
amide-alpha-carbon bond 300
Amino Acid Modification Card (AAMC) 329
amino acids
  adding to chemical modifications 201
  Amino Acid Modification Card (AAMC) 329
  charge given to proteins 244
  deleting from chemical modifications 204
  displaying post-translational modifications (PTMs) 319, 323
  in Chemical Modifications view 197
  in Peptide Consensus view 322
  in Protein Identification Details view 248, 256
  in Sequence Comparison view 284
  mass values 711
  number found during FASTA file processing 161
  symbols 711
  unmodified 323
annotated spectra 3, 578, 582
Annotation column 299
Annotation node
  description 16, 120, 704
  retrieving protein annotations 249, 253, 433, 437
Annotation view 436
annotation workflow 437
Annotation workflow nodes 703
antioxidant activity GO Slim category 459
Any filter 98
application menus 22
Apply icon 199
Area column 244, 295
arginine 470
Auto Hide icon 34, 259, 272, 278
Auto Layout command/icon 26, 625
Auto Number command/icon 626
available search engines 3

B

b* fragment ions 713–714
b° fragment ions 713–714
b fragment ions
  definition 7, 302, 714
  displayed in Peptide Consensus view 317
  displayed in Peptide Identification Details view 302, 311, 314
  flag representing 318
  generated by CID 3

b fragment ions (continued)
location of charge 300
used for spectrum matching 676
base peak chromatogram
in extracted ion chromatogram view 340
Rawfile and Scan Range Selection page of search wizards 72
batch processing
creating workflow 123–124
definition 124
in Discoverer Daemon 121, 124–125
monitoring job execution in Discoverer Daemon 127
reannotating MSF files 444
selecting single sample 139
setting up experiment method with Sequence Editor 136
setting up mass spectrometer 136
using a post-acquisition method 144
using a processing method 134
biological process codes 457
Biological Process column 243, 437
Biological Processes page 449, 456
BioWorks
converting 3.3 or earlier MS^n data to Proteome Discoverer formats 726
converting 3.3.1 MS^n data to Proteome Discoverer formats 715, 721
failed migrations to Proteome Discoverer application 737
installation directories created in 716
installing for data migration 717
BioWorks 3.3.1 SP1 setup dialog box 718
BioWorks Browser Setup wizard 717–718
BioWorksBrowser prompt box 732
Blast searches 705
BMP format 577, 598, 637

C

c^o fragment ions 713–714
c fragment ions
abstracting proton from precursor 7
definition 7, 302, 714
displayed in Peptide Consensus view 317
displayed in Peptide Identification Details view 302, 311, 314
flag representing 318
generated by ETD and ECD 3
location of charge 300
used for spectrum matching 676
C terminus 80–81, 88
calc. pI column 244
Cancel icon 161
catalytic activity GO Slim category 459
cell communication GO Slim category 463
cell death GO Slim category 463
cell differentiation GO Slim category 464
cell division GO Slim category 464
cell growth GO Slim category 464
cell homeostasis GO Slim category 464
cell motility GO Slim category 464
cell organization and biogenesis GO Slim category 464
cell proliferation GO Slim category 464
cell surface GO Slim category 460
cellular component codes 456
Cellular Component column 243
Cellular Components page 449, 455
Centroid Sum integration method 480, 701
Centroid With Smallest Delta Mass integration method 480, 701
Charge column 295
Charge State peptide filter
Mascot reports 371
SEQUEST reports 366
ZCore reports 375
charge-reduced precursor peaks 649, 651, 654
Check All command 631, 634, 636
Check Modifications page 722
Check Selected command 631, 634, 636
chemical modifications
adding 200
adding amino acids 201
deleting 202
deleting amino acids 204
displaying 196
dynamic 196
importing 202
importing from local file 203
importing from UNIMOD 202
selecting in SRF File Import Wizard 722
static 196
updating existing 201
Chemical Modifications view
Abbreviation column 199
adding modifications 200
Apply icon 199
Delta Average Mass column 199
Delta Mass column 199
description 195–196
Import icon 199
Is Active column 199
Leaving Group column 199
Modification column 199
opening 196, 627
parameters 199

Chemical Modifications view (continued)
Position column 200–201
Remove icon 199
Substitution column 199
Unimod Accession No. column 200
Chromatogram view
copying 333
description 330
displaying 330
Extracted Ion Chromatogram 331
features 331
opening 620
parameters 333
saving in image format 333
zooming in and out 332
chromosome GO Slim category 460
CID activation type
analyzed by SEQUEST 3
description 8
fragmenting ions 7, 300, 713
generating Dual-Peptide Consensus view 319
ion factors 78
selecting as peptide filter
in Mascot reports 371
in SEQUEST reports 366
in ZCore reports 375
selecting in search wizards 74, 85
specifying
in Reporter Ions Quantifier node 481, 700
in Scan Event Filter node 660
in Spectrum Selector node 640, 645
Clear All Filters icon 178
Clear Filter Criteria icon 395
Clear Reference Filter icon 178
cleavage properties 712
cleavage reagents
adding 211
cleavage properties 712
deleting 211
displaying 206
filtering data 212
modifying 211
specifying in Mascot 666
specifying in SEQUEST 672
specifying in ZCore 678
Cleavage Reagents view
description 206
opening 206, 627
parameters on 207
cleavage specificities 210
Close All Windows command 629
Close command 616
coagulation GO Slim category 464
coisolation 513, 547
Color dialog box 55
colors in interface 21
Column Chooser dialog box
organizing reports 223
Peptides page 291
Proteins page 241
Search Input page 347
Column Chooser icon 223
CombIon 737
Commands page 49
Compact icon 161, 163
Compile FASTA Database page 170, 174–175
Complete the Wizard page 725
Completing the *Wizard_name* Search Wizard page 81, 89
compressing protein databases 163
Confidence icon column 291
confidence indicators
applying filter settings to change 429
distribution in target false discovery rates 416
exporting to Excel spreadsheet 607
in peptide rows on Proteins page 239
in Protein Identification Details view 247
in Sequence Comparison view 284–285
on Peptides page 416
Configuration command/icon 26, 627
Configuration page 129
Configuration view
configuring protein annotation 436
configuring search engines 64
opening 627
configuring Mascot 66–67
configuring protein annotation 435
configuring SEQUEST 65
configuring ZCore 70
conjugation GO Slim category 464
Copy command
graphs 637
Peptides page 633
Proteins page 631
Search Input page 635
Copy With Column Headers command 631, 633, 635
Coverage column 242
Coverage page 246
Create Bug Report command 628
Create Quantification Method dialog box 515, 517, 519
Create SRF File dialog box 731
Creator ID column 347
cSNPs 706
Custom Filter dialog box 397, 399
Custom filter dialog box 212

custom parsing rule A 708
custom parsing rule B 709
custom parsing rule C 709
Customize dialog box
  adding new toolbar 39
  customizing toolbar features 41
  deleting toolbar 40
  opening 37
  parameters on 48
  removing toolbars 38
  renaming toolbars 40
Customize Keyboard dialog box
  adding a shortcut key 45
  parameters on 50
  removing a shortcut key 46
Customize menu 20
cytoplasm GO Slim category 460
cytoskeleton GO Slim category 461
cytosol GO Slim category 461

D

Data Archive files *See* DTA files
Data Export workflow nodes 702
Data Input nodes 638
deconvolution 546
decoy database search
  calculating false discovery rates 415
  Peptide Validator 681
  Percolator 681
defense response GO Slim category 464
Delete Methods dialog box 521
Deleted FASTA Indexes table 191
Delta Average Mass column 199
Delta Cn column 294, 386
Delta Cn value 363, 385–386, 409–410, 690
Delta M column 298
Delta Mass column 199
delta retention time 343
delta score 345
Delta Score column 294, 345
Description column 241
development GO Slim category 465
dialog box icons 35
Differentiable Proteins protein filter 364
dimethylation 3plex quantification method
  description 469, 472
  selecting in Quantification Editor dialog box 488, 490, 507
  troubleshooting precursor ion quantification 575
Diseases page 449, 458
Display Temporary icon 161
Distinct Proteins protein filter 364
.dlyt suffix 235
DNA 459–460
DNA binding GO Slim category 459
documentation survey xviii
DotNet installation 718
down arrow icon 395
DTA files
  corrupted 738
  creating from DTA and FASTA files 730
  creating SRF files 728
  exporting spectra through Spectrum Exporter node 114
  importing into Proteome Discoverer application from BioWorks 730
  moving to BioWorks 3.3.1 731
  output file type 13, 577–578, 580, 703
  role in failed data migrations 737
Dual-Peptide Consensus view
  confirming peptide identity 319
  description 324
  displaying 324
  features of 326
dynamic modifications
  definition 196
  selecting in SEQUEST wizard 79
  setting in Mascot 666
  setting in SEQUEST 671
  setting in ZCore 678

E

EC number 455–457
ECD activation type
  analyzed by SEQUEST 3
  description 8
  fragmenting ions 300, 713
  ion factors 78
  selecting as peptide filter
    in Mascot reports 371
    in SEQUEST reports 366
    in ZCore reports 375
  selecting in search wizards 74, 85
  specifying
    in Reporter Ions Quantifier node 481, 700
    in Scan Event Filter node 660
    in Spectrum Selector node 640, 645
ECD spectra 649
Edit Administration icon 715
Edit Quantification Method command/icon 25, 520, 622
elution time of peptides 341, 343
elution time of proteins 333
EMF format 577, 598, 637

Enable Protein Grouping command 402, 630, 633, 635
Enable Row Filters command 632, 634, 636
endoplasmatic reticulum GO Slim category 461
endosome GO Slim category 461
Ensembl database 452, 706
Enter filter criteria for Reference dialog box 168, 179
Entrez Gene Web site 451
Enzyme Category (EC) information 454
enzyme regulator activity GO Slim category 459
enzymes 712
ESI
  fragment ions 713
  peptides and fragment ions 7
  with PQD 9
ETD activation type
  analyzed by SEQUEST 3
  analyzed by ZCore 4
  description 8
  fragmenting ions 300, 713
  generating Dual-Peptide Consensus view 319
  ion factors 78
  selecting as peptide filter
    in Mascot reports 371
    in SEQUEST reports 366
    in ZCore reports 375
  selecting in search wizards 74, 85
  specifying
    in Reporter Ions Quantifier node 481, 700
    in Scan Event Filter node 660
    in Spectrum Selector node 640, 645
ETD spectra
  Non-Fragment Filter node 649
  ZCore node 677
ETD Spectrum Charger node
  description 117, 648
  known issues 741
European Bioinformatics Institute (EBI) 454
Event Detector node
  description 117, 639
  parameters 639
  peak area calculation quantification 483–484
  precursor ion quantification 473–474
evidence codes 454, 456–457
Example Workflow Templates folder 718
Excel
  2007 604
  97-2003 604
  exporting data from a TXT file to 612
  exporting Modfication List page to 254
  exporting search results to 232, 236, 577, 604–605
Excluded by Method peptide status 297
exclusion mass lists
  definition 600
  export options 577
  exporting to Xcalibur data system 600
Exit command 618
expectation value 294, 374
Expectation Value column 294
Experimental Bias page 501
Export Annotated Spectra dialog box 582, 594, 617
Export command 617
Export Exclusion/Inclusion List dialog box 600, 603, 618
Export Parameter File page 126, 149
Export Spectra dialog box 578–579, 617
Export to Excel Workbook command 631, 636, 633
Export to Excel Workbook dialog box 605
Export to FASTA dialog box 599–600, 609, 618
Export to ProteinCenter command 628
Export to ProtXML dialog box 596, 618
Export Workflow to XML command 626
Extensible Markup Language files. *See* XML files
External Links page 449, 458
extracellular GO Slim category 461
Extracted Ion Chromatogram view
  accessing from Chromatogram view 331
  description 340
  displaying 341
  features 340
  opening 621
  parameters 343

F

false discovery rate
  definition 415
  recalculating 426
  relaxed
    peptide confidence indicators 416
    Peptide Confidence page 424
    Peptide Validator node 681
    Percolator node 690
    specifying for decoy database search 416
  setting up in search wizards 418
  setting up in Workflow Editor
    Peptide Validator node 419, 681
    Percolator node 421, 682
  strict
    peptide confidence indicators 416
    Peptide Confidence page 424
    Peptide Validator node 681
    Percolator node 690
    specifying for decoy database search 416
  target 416

FASTA Database Utilities command 628
FASTA Database Utilities dialog box
  Add Protein References page 165, 175
  Compile FASTA Database page 170, 174–175
  Find Protein References page 165, 177
  opening 164, 628
  parameters on 175
FASTA files
  adding 161–162, 626, 727
  adding before using search wizards 71
  adding protein reference 164
  adding protein sequence 164
  cancelling addition or removal 161
  compiling 170
  creating SRF files 722, 728
  deleting 162
  displaying 159
    current status 161
    date last modified 161
    name 161
    number of amino acids found 161
    number of sequences found 161
    size 161
    temporary 164
  excluding protein sequences or references 174
  exporting protein search results to 618
  exporting proteins 599
  exporting search results to 232
  filtering protein reference searches 167
  finding protein sequences or references 165
  importing SRF files 721
  input to search wizards 63
  modifying 626
  most important databases 705
  parsing rules 708
  removing 626
  selecting in SRF File Import Wizard 722
  selecting in Workflow Editor 728
  updating to match SRF files 734
FASTA files view
  # Residues column 161
  # Sequences column 161
  Cancel icon 161
  Compact icon 161, 163
  Display Temporary icon 161
  Last Modified column 161
  Name column 161
  opening 159, 626
  parameters on 161
  Size column 161
  Status column 161
FASTA Index Creator dialog box 185, 193
FASTA indexes
  adding 626
  automatic removal 190
  automatically creating 185
  changing maximum number stored 191
  changing storage location 191
  deactivating automatic removal 190
  definition 179
  deleting 190
  deleting from deleted FASTA files 193
  discarding changes from previous session 193
  displaying 181
  manually creating 17, 189
  removing 626
  restoring deleted 191
FASTA Indexes Options dialog box 192, 195
FASTA Indexes view
  after adding FASTA index 188–189
  opening 181
  parameters on 183
FDR. *See* false discovery rate
Features page 449, 453
File menu 22, 616
File Selection page 721
filter sets
  copying from one installation of Proteome Discoverer to another 389
  creating 389
  deleting 391
  loading 389–390
  saving 389
  using 389
Filtered Out peptide status 298
filtered proteins compared to loaded proteins 27
filtered search inputs compared to loaded search inputs 27
filtered-out rows 400
filters
  deactivating 392
  false discovery rates 415
  peptide 361
  protein 361, 364
  protein reference searches 167
  removing 392
  Result Filters page 361
Find Protein References page 165, 177
First Scan column 298, 348
Fourier Transform mass spectrometer 480, 661, 642
  specifying in Reporter Ions Quantifier node 702
  specifying in Spectrum Selector node 646
fragment ions
  activation types producing 713
  ammonia loss 713

fragment ions (continued)
changing display units 318
charge state 319
charged on C-terminal side 713
charged on N-terminal side 713
display units 313
displaying in Peptide Identification Details view 302, 308
displaying masses 323
exporting spectra annotated with 586–587
factors dependent on 7
identifying 300
immonium ions 58, 302–303, 311
neutral loss ions 58, 302–303, 311
phosphorylation losses 302–303
post-translational modifications (PTMs) 58, 302–303, 312
precursor ions 58, 302–303, 311
resulting from CID and HCD activation 300
resulting from ETD and ECD activation 300
selecting to display 318
setting tolerance between measured and theoretical 319
symbols representing 318
types 7, 713
units of display 323
water loss 713
Fragment Match Colors and Fonts page 54, 59
fragment match options 53, 301
Fragment Match Options dialog box 585
Fragment Match Options page 53, 57
Fragment Match Spectrum view
description 335
displaying 335
opening 620
parameters on 337
Fragment Matches pane 301
Fragment Spectrum pane 301
fragmentation methods. *See* activation types
FT ICR mass spectrometers 650, 657
full scan type 644, 662
funnel icon 212

G

gel electrophoresis 243–244
GenBank database 705, 708
Gene Ontology (GO) database *See* GO database
General page 449–450, 504
GIF format 577, 598, 637
global default fragment match options 53
GO codes 454, 456–457
GO database
description 434
features 435
retrieving information from ProteinCenter 433, 438
Web site 435
GO Slim categories
biological processes 463
cellular components 460
colors 439
definition 455–457
molecular components 459
GO terms 434, 455
golgi GO Slim category 462
Graph shortcut menu 636
Group 242
Group Scan 737

H

Has Modifications peptide filter
Mascot reports 372
SEQUEST reports 367
ZCore reports 376
HCD activation type
analyzed by SEQUEST 3
description 8
fragmenting ions 300, 713
grouping scans 656
ion factors 78
selecting as peptide filter
in Mascot reports 371
in SEQUEST reports 366
in ZCore reports 375
selecting in search wizards 74, 85
specifying
in Reporter Ions Quantifier node 571, 700
in Scan Event Filter node 660
in Spectrum Selector node 640, 645
Heavy/Light column 244, 298
Heavy/Light Count column 244, 560
Heavy/Light Variability column 244, 560
Help Contents command 629
Help menu 22, 629
Hidden Markov Model 243, 435
high-confidence peptides
Peptides page 291
protein sequence coverage bar 256
highest-scoring proteins 221
homologous proteins 363, 404
Homology Threshold column 296
how Protein Group Members View command 402

HTML files
annotated spectra for peptides belonging to selected proteins 589
annotated spectra for selected peptides 589
annotated spectra of peptides of selected peptide groups 590
annotated spectrum example 592
compressing 587
contents of exported 3, 582, 588
index page 584
peak list 593
peptides table showing annotated spectra for peptides 588
search summary 591
Human Proteome Organization (HUPO) 12–13, 115, 579–580, 703

I

ID column 347
Identified Peptides column 347
Identity High column 295
Identity Middle column 296
immonium ions 58, 302–303, 311
Import icon 199
Import Modifications dialog box 203
Import Search Results from SRF command 222, 616
Import Workflow dialog box 112
importing 3.3 or earlier MS^n data from BioWorks 726
importing chemical modifications 202
importing MS^n data from BioWorks 715, 721
Include/Exclude Peptide(s) From Protein Quantification command 632, 634, 636
Inconsistently Labeled peptide status 298
incorrect node parameters 110
Indistinguishable Channels peptide status 297
infrared multi-photon dissociation. *See* IRMPD activation type
Input Files page
combining results of multiple searches 351
moving files 217
opening a report file 215
parameters on 219
removing files from 216, 219
inputs to Proteome Discoverer application 12
installation directories 716
Installation Guide command 629
installing workflow template 718
Instrument Setup icon 136
Intensity column 298
International Union of Biochemistry and Molecular Biology 455–457
InterPro database 453
Ion Inject time column 295, 348
Ion Series page 302
ion trap mass spectrometer 480
processing PQD data from 571–572
specifying in Reporter Ions Quantifier node 702
specifying in Scan Event Filter node 661
specifying in Spectrum Selector node 642, 646
ionization sources 641, 660
Ions Matched column 299
IonScore column 294
IonScore value
combining filters in multiconsensus reports 357
determining Identity High parameter value 295
on Peptide Identification Details view 312
Peptides page 233, 294, 353, 683
IPI (International Protein Index) database 706
IRMPD activation type 9, 713
Is Active column 199
Is filter 98
Is Master Protein column 245
Is Not filter 98
isobaric tags for relative and absolute quantification. *See* iTRAQ quantification
isoelectric focusing 244
isoelectric point 244, 262
isoforms
effect of too many on pRS site probabilities calculation 692
phosphorylation 695–696
probability of being correct 299, 691
pRS score of 696
putative 694, 697
% Isolation Interference column 294
isoleucine 363, 404
isotope intensity 494, 510
isotope patterns 569
isotope shift 494, 510
isotopomers 469
iterative searches 16, 664
iTRAQ 4plex quantification method
as default 7, 477, 488
selecting in Quantification Method Editor dialog box 493, 509
iTRAQ 8plex quantification method
as default 7, 477, 488
selecting in Quantification Method Editor dialog box 493, 509

iTRAQ quantification
creating a workflow for 477
default methods available in 477, 488, 493, 509
description 7, 477
isotopic distribution values 545
performing 476
Reporter Ions Quantifier node 7, 120
specifying label modifications 489
See also reporter ion quantification
IUBMB Enzyme Nomenclature 455–457

J

job queue
deleting searches 32
description 28
displaying 26, 28, 626
in creating FASTA indexes 187, 189
opening 71
opening completed report 33
opening MSF file 219
parameters 30
Search ID command 241, 292
toolbars
adding 51
customizing 51
deleting 53
displaying and hiding 51
renaming 52
See also Discoverer Daemon
Job Queue page 127
JPG format 577, 598, 637

K

Keys page 449, 452

L

Last Modified column 161
Last Scan column 298, 348
Layout command 621
layouts
applying default to report 235
applying new to report 235
automatically aligning 625
creating 234
creating custom 233
creating default 235
limitations of custom 233
returning to original 236
saving 234, 236
LC-MS 9
LC-MS/MS 4
Leaving Group column 199
leucine 363, 404
Licenses view 627
limitations 14
Load Filter Set dialog box 390
logic operator and command menu icon 395
logic operator menu icon 395
LTQ FT mass spectrometers
ETD Spectrum Charger node 648
Non-Fragment Filter node 654
Spectrum Selector node 642
LTQ Orbitrap mass spectrometers
adding a non-fragment filter node 106
ETD Spectrum Charger node 648
mass analyzer 642
non-fragment filter node 650
precursor mass criterion 657
troubleshooting 571
using ZCore search wizard 63
workflow demonstration 99
LTQ XL mass spectrometers 63
lysine 470

M

m/z column 298
Magellan server 122, 130
Magellan storage files. *See* MSF files
Maintain Chemical Modifications
command/icon 26, 196, 627
Maintain Cleavage Reagents command/icon 26, 206, 627
Maintain FASTA Files command/icon 26, 626
Maintain FASTA Files icon
adding FASTA files 162
listing FASTA files 159
Maintain FASTA Indexes command/icon 26, 626
Maintain Quantification Methods command/icon
adding quantification method 516
changing quantification method 520
deactivating quantification method 522
exporting quantification method 523
importing quantification method 522
opening Quantification Methods view 26, 627
removing quantification method 521
restoring quantification method template defaults 503
setting up Quantification Method Editor dialog box 486
MALDI 9
Manage Licenses command/icon 26, 627
Mascot Communication Tester command 628
Mascot Generic Format files. *See* MGF files

Mascot node
  description 119, 666
  parameters 666
Mascot search engine
  configuring 66–67
  description 3–4
  directing application to server location 67
  highest-scoring peptide 224
  IonScore value
    combining filters in multiconsensus reports 357
    custom layouts 233
    determining Identity High parameter value 295
    Peptide Identification Details view 312
    Peptides page 294, 353, 683
  main score 345
  options for calculating FDR 426, 429
  peptide filters 371
  starting 71
  troubleshooting failed searches 69
  wizard 625
Mascot Significance Threshold peptide filter
  applying to search reports 356
  Mascot reports 372
  Peptide Confidence page 429
  recalculating false discovery rates 426
Mass Analyzer column 347
mass tags 493, 495, 509
mass tolerance unit 224
Mass Tolerance Unit command 631, 633, 635
master proteins 402
Master Scan(s) column 348
membrane GO Slim category 462
menus in the application 22
metabolic process GO Slim category 465
metal ion binding GO Slim category 459
MGF files
  contents of 68
  importing into Workflow Editor 114
  input file type 12, 63
  output file type 13, 114, 577–578, 580, 703
  splitting 68
MH+ column 298
Missing Channels column 295
missing reporter ions 535
mitochondrion GO Slim category 462
.mlyt suffix 235
Modification column 199
Modification List page 251
Modifications column 293
Modifications page, parameters on 88
molecular function codes 455
Molecular Function column 243
Molecular Functions page 449, 454
Most Confident Centroid integration method 480, 701
Most Intense Centroid integration method 480, 701
motor activity GO Slim category 459
MPD activation type
  selecting as peptide filter
    in Mascot reports 371
    in SEQUEST reports 366
    in ZCore reports 375
  selecting in search wizards 74, 85
  specifying
    in Reporter Ions Quantifier node 481, 700
    in Scan Event Filter node 660
    in Spectrum Selector node 640, 645
MS Order column 299, 347
MS spectra
  Peptides page 299
  Precursor Isotope Pattern view 338
  Search Input page 347
MS/MS spectra
  displaying in Peptide Consensus view 322
  displaying number of scan 348
  excluding mass-to-charge ratios 600, 618
  fragmenting reporter ions in 529
  generating CID 7, 713
  grouping 118
  minimum ion count 75
  Minimum Ion Count command 85
  order 299, 347
  portion derived from MS scan in Isotope Pattern view 338
  processed by Mascot 382
  processed by SEQUEST 382
  processed by ZCore 4
  removing artifacts with Non-Fragment Filter node 649, 655
  reporter ion quantification 477
  SEQUEST processing 4
  signal-to-noise ratios 341
  sorting 656
  triggering for same precursor 224
  types of fragment ions observed in 7
  using for two separate peptide matches 297
  validating searches with FDRs 415
  viewing in Fragment Match Spectrum view 620
  viewing in Spectrum view 620
  when taken in Extracted Ion Chromatogram 340, 343
MS3 spectra 299, 347
MSDB database 708

MSF files
activating Quantification menu 468
adding fixed row index numbers 231
adding row index numbers to visible and invisible rows 231
adding row index numbers to visible rows only 230
applying default layouts 235
applying new layouts 235
choosing columns to display 223
closing 222
creating custom layouts 233
creating default layout 235
customizing data 222
displaying highest-scoring peptides 224
displaying peptides included in proteins 225
displaying peptides on Search Input page 225
displaying Peptides page 288
displaying Proteins page 238
example 214
exporting data to Microsoft Excel 236
filtering data 379
filtering data before opening 218
grouping data by column header 227
groupingpeptides data by sequence 228
input to Proteome Discoverer application 112
limiting number of proteins displayed 218
multiple files in quantification 503
opening
from job queue 219
from Result Filters page 218
through menu bar or toolbar 215
output by Proteome Discoverer application 13
pages in 214
reannotating 443–444
removing row index numbers 232
returning to original layouts 236
saving 221
saving layout 236
selecting peptides and proteins 232
sorting columns 229
specifying mass tolerance unit 224
treating quantification results as replicates 217, 220
ungrouping proteins 406
MSIPI database 706
MudPIT
creating a search workflow 101
creating parameter file to call Discoverer Daemon from Xcalibur data system 126
creating workflow for multiple .raw files 101
creating workflow for processing 123–124
description 9, 124
monitoring job execution in Discoverer Daemon 127
processing samples in Discoverer Daemon 121, 125, 154
scoring 670
Start Jobs page in Discoverer Daemon 124
using processing method 149
using Run Sequence dialog box 153
when to use to search for sample fractions 101
multiconsensus reports
applying layouts 233
calculating and displaying protein ratios in mixed mode 567
calculating and displaying protein ratios in reports treated as replicates 563, 566
calculating and displaying protein ratios in reports treated as treatments 563, 565
calculating variability in reports treated as replicates 563
calculating variability in reports treated as treatments 562
combining results of multiple searches 351
description 351, 353
example 352
filtering 356
generated by more than one search engine 233
Peptides page 353
Proteins page 353
setting up in Quantification Method Editor dialog box 566
suffix 235
treated as replicates 561
treated as treatments 561
when to use to search for sample fractions 101
with quantification results 356
Multidimensional Protein Identification Technology. *See* MudPIT
multiple MSF files 503
multiple search results 213
multiple searches 121, 351
MW column 243
MZDATA files
importing into Workflow Editor 114
input file type 12, 63
output file type 13, 114–115, 577–580, 703
MZML files
input file type 12, 63, 114
output file type 13, 15, 115, 577–580, 703
MZXML files
input file type 12, 63, 114
output file type 13, 115

N

N terminus 80–81, 88
Name column 161

National Center for Biotechnology Information (NCBI) 451, 705
National Institutes of Health (NIH) 455–457
NCBI RefSeq database 706
neutral loss ions
  automatic color assignment 54, 59, 302
  identifying in grouped scans 309
  in Peptide Identification Details view 306, 311
  removing 107
  selecting for annotation in Options dialog box 58
  specifying default global colors and fonts 53
  types of 303
neutral loss peaks 107, 651, 653–654
Neutral Losses page 307
New Sequence Template dialog box 142
New Toolbar dialog box 39, 52
New Workflow command/icon 26, 625
No proteins peptide status 297
No Quan Labels peptide status 297
No Quan Values peptide status 297
Noise Peak Filter node
  description 117, 648
  parameters 649
Non-Fragment Filter node
  adding for high-resolution data 106
  calculating mass window for charged-reduced precusor peaks 651
  calculating mass window for neutral loss peaks 651
  calculating mass window for precursor overtone peaks 651
  calculating mass window for precursor peaks 650
  description 118, 649
  parameters 653
non-fragment ion peaks 649
normalization factors 502, 514
Not Unique peptide status 297
NRL3D database 708
nucleotide binding GO Slim category 459
nucleus GO Slim category 462
AAs column 243
Proteins column 242, 402
Unique Peptides column 242, 402, 412
Missed Cleavages column 294

O

^{18}O labeling quantification method
  description 469, 472
  selecting in Quantification Method Editor dialog box 488, 490, 507
  troubleshooting precursor ion quantification 575
Open command/icon 216, 218, 352
Open From Template command/icon 26, 625
Open Processing Workflow Templates dialog box 109, 112–113
Open QualBrowser command/icon 205, 628
Open Report command/icon 24, 215, 351, 616
Open Report icon 51, 219, 726, 737
Open Search Results File dialog box 727
Operation done page 724
Operation status page 724
Options command 628
Options dialog box
  description 302
  Fragment Match Colors and Fonts page 54, 59
  Fragment Match Options page 53, 57
  opening 628
  parameters 56
  ProteinCenter page 448
Options page 41, 47–49
Options pane 301
organelle lumen GO Slim category 463
OUT files
  corrupted 738
  creating SRF files 728
  importing into Proteome Discoverer application from BioWorks 730
  moving to BioWorks 3.3.1 731
outdated workflow nodes 109
outputs of Proteome Discoverer application 13
overtones 107

P

parallel workflows 105
parameter file 125, 129, 132, 149
Parameters pane 6, 96
Pause icon 51
PDB database 705
peak area calculation quantification
  creating workflow for 483
  description 7, 483
  performing 483
  Precursor Ions Area Detector node 483
pen icon 113
PEP *See* posterior error probability (PEP)
Peptide Confidence page
  changing filter settings 427
  changing target rate 427
  functions 425
  parameters on 429
  recalculating false discovery rate 426
  viewing decoy database search results 424

Peptide Confidence peptide filter
  Mascot reports 372
  SEQUEST reports 367
  ZCore reports 376
Peptide Consensus view
  accessing UNIMOD Web view 327
  changing units to display fragment ion masses 318
  controlling visibility of ion table 318
  description 316
  displaying 316
  displaying different fragment ions 318
  displaying fragment ions by charge state 319
  displaying information on post-translational modifications (PTMs) 319
  features 317
  opening 619
  parameters 321
  setting tolerance between measured and theoretical ion fragments 319
Peptide Delta Cn peptide filter
  Mascot reports 373
  SEQUEST reports 367
  ZCore reports 376
Peptide Delta Cn value 367, 373, 376
peptide filters
  applying 380
  Delta Cn 385
  Mascot 371
  rank 382, 385
  search engine rank 387
  SEQUEST 366
  ZCore 375
Peptide Group Members view
  columns in 691
  contents when peptide grouping not enabled 344
  description 343
  displaying 343, 620
  pRS Probability column 691
  pRS Score column 691
  pRS Site Probabilities column 691
Peptide Identification Details dialog box 53
Peptide Identification Details view
  description 299
  displaying 300
  displaying fragment ions 302, 308
  displaying neutral loss ions 306
  displaying precursor ions 307
  features 301
  Fragment Matches pane 301
  Fragment Spectrum pane 301
  grouping scans 309
  ion series colors 311
  Ion Series page 302
  Neutral Losses page 307
  opening 619
  Options pane 301
  parameters on 310
  Peptide Summary pane 301
  Precursor Ions page 307
Peptide Length peptide filter
  Mascot reports 373
  SEQUEST reports 368
  ZCore reports 377
Peptide Mass Deviation peptide filter
  Mascot reports 373
  SEQUEST reports 368
  ZCore reports 377
Peptide Rank peptide filter
  Mascot reports 373
  SEQUEST reports 368
  ZCore reports 377
Peptide Ratio Distributions chart 550, 552, 622
peptide ratio variability 244
peptide ratios
  calculating 550, 552
  deriving protein ratios from 557
  displaying number used to calculate protein ratios 500
  handling missing and extreme values 554
  setting up 497
Peptide Score peptide filter
  Mascot reports 356, 374, 426
  Peptide Confidence page
    Filter Settings 431
    selecting a filter 429
  SEQUEST reports 369, 426
  ZCore reports 378, 426, 429
peptide spectrum matches. *See* peptides
Peptide Summary pane 301
Peptide Validator node
  attaching to search engine nodes 94
  description 16, 119, 680
  parameters 419, 680
  setting up false discovery rates in Workflow Editor 419
Peptide/Protein ID workflow nodes 666
Peptide/Protein Validation nodes 680
Peptides column 242
peptides
  C terminus 80–81, 88
  calculating ratios 550, 552
  Chromatogram view 330
  classifying for protein quantification 548
  confidence indicators 416, 607
  defining uniqueness 469

peptides (continued)
displaying
filtered-out 254–255, 630, 633, 635
filtered-out rows 400
highest-scoring in MSF file 224
those included in proteins 225
displaying on Search Input page 225
Dual-Peptide Consensus view 324
excluding from protein quantification 546
excluding those with high levels of coisolation 547
expanding identified 402
exporting annotated spectra of 582–583, 594
Extracted Ion Chromatogram view 340
filtering 361, 380
by rank 382, 385
deactiving filters 392
Mascot filters 371
removing 392
result filters 379
row filters 393
SEQUEST filters 366
filtering by Delta Cn 385
filtering by search engine rank 387
Fragment Match Spectrum view 335
grouped on Peptides page 290
grouping
on Peptides or Proteins page shortcut menu 413, 630
on Result Filters page 362
on Results Filters page 413
on Search Input page shortcut menu 635
on the Result Filters page 413
options 414
grouping by sequence 228
grouping data by column header 227
grouping on Peptides or Proteins page shortcut menu 632
high levels of co-isolation 548
high-confidence
Peptide Validator node 420
Peptides Confidence page 431
Peptides page 22
Percolator node 422
Protein Identification Details view 247, 256
Proteins page 239
Result Filters page 363, 405
search wizards 419
Sequence Comparison view 284–285
Spectrum Confidence Filter node 665
low-confidence
Peptides page 291
peptides page 22
Protein Identification Details view 247, 256
Proteins page 239
Result Filters page 363, 405
Sequence Comparison view 284–285
Spectrum Confidence Filter node 665
medium-confidence
Peptide Validator node 420
Peptides page 22, 291
Percolator node 422
Protein Identification Details view 247, 256
Result Filters page 363, 405
search wizards 419
Sequence Commparison view 284
Sequence Comparison view 285
Spectrum Confidence Filter node 665
N terminus 80–81, 88
neutral loss ions 309
number displayed on status bar 413
number of filtered compared to loaded 27
Peptide Consensus view 316
Peptide Identification Details view 299
precursor ions 337
Precursor Isotope Pattern view 337
selecting 232
Spectrum view 333
ungrouped on Peptides page 290
ZCore filters 375
Peptides Containing Any of the Specified Amino Acids peptide filter
Mascot reports 374
SEQUEST reports 369
ZCore reports 378
Peptides Matched column 295
Peptides page
adding fixed row index numbers 231
adding row index numbers to visible and invisible rows 231
Chromatogram view 330
creating custom layouts 233
description 288
displaying 288, 630
displaying filtered-out rows 400
Dual-Peptide Consensus view 324
exporting search results to Microsoft Excel 236
exporting to tab-delimited TXT file 15, 609
Extracted Ion Chromatogram view 340
features 289
Fragment Match Spectrum view 335
grouped peptides 290
multiconsensus reports 353
numbering rows 230
parameters 291
Peptide Consensus view 316
Peptide Identification Details view 299
Precursor Isotope Pattern view 337
Quan Info column colors 22

Peptides page (continued)
Ratio columns 495, 497
removing row index numbers 232
row filters 393
selecting peptides and proteins 232
Spectrum view 333
ungrouped peptides 290
Peptides page of the To CSV dialog box 611
Peptides page shortcut menu 632
Peptides Per Protein protein filter 364
pepXML files 13, 597
% Isolation Interference column 348
Percolator filter 426, 429
Percolator node
algorithm used 682
attaching to search engine nodes 94
description 16, 119, 682
feature sets used 683
parameters 690
posterior error probability (PEP) 688
q-value 688, 690
setting thresholds for scores 425
setting up false discovery rates in Workflow Editor 421
very small searches 421, 682
weights assigned to features in feature set 686
Pfam database
accession identifier 454
computational enrichment 453
description 435
features 435
Hidden Markov Model 243, 435
Pfam IDs column 243, 435
phospho*RS* node
calculating pRS score 691, 696
calculating pRS sequence probability 691, 696–697
calculating pRS site probabilities 691, 694, 696
creating PTM analysis workflow 103
description 16, 120, 691
effect of too many isoforms on pRS site probabilities calculation 692
internal workflow 694
making PTM site probabilities available in Protein ID Details view 693
parameters 698
PTM site localization scores in Protein ID Details view 249, 255
type of grouped data to use with 693
phosphorylation
calculating pRS site probabilities for 691, 697
calculating PTM site localization scores with phospho*RS* node 103
color-coding in Protein ID Details view 249
isoforms 695
pRS Site Probabilities column on Peptides page 299
PTM site localization scores in Protein ID Details view 249, 255
phosphorylation losses 302–303, 311
pie charts
copying 279
displaying 278
parameters for 280
saving 279
PIR database 705, 708
Please Select a FASTA Database dialog box 166
PMD files 136
PNG format 577, 598, 637
Position column 200–201
post-acquisition method 144
posterior error probability (PEP) 688, 690
post-translational modifications (PTMs)
analyzing with Peptide Consensus view 316
annotating spectra with 58
determined by dynamic modifications 196, 671
displaying 253–255
displaying in Protein Identification Details view 245
factory default settings for mass analyzer/activation type combination 53
fragment match colors of 54, 59, 302
how denoted in fragment ion notation 329
in Peptide Identification Details view 303, 312
molecular weight on Proteins page 243
site localization scores 245–246, 249
site probabilities 249, 255
viewing information 319
PQD activation type
description 9
ion factors 78
selecting as peptide filter
in Mascot reports 371
in SEQUEST reports 366
in ZCore reports 375
selecting in search wizards 74, 85
specifying
in Reporter Ions Quantifier node 481, 571, 700
in Scan Event Filter node 660
in Spectrum Selector node 640, 645
precursor artifact peaks
Mass Window Offset option 653
Remove Precursor Peak option 655
removed by Non-Fragment Filter node 649
Precursor Charge column 348
Precursor Intensity column 348

precursor ion quantification
  calculating peptide ratios 550, 554
  checking the quantification method 502
  columns on Peptides page unique to 295
  columns on Proteins page unique to 244
  controlling protein and peptide ratios 497
  correcting experimental bias 501
  correction for isotopic impurities 498
  creating workflow for 472
  default methods available in 469, 487, 490, 507
  description 469
  displaying quantification channel values 529–530
  displaying Quantification Spectrum chart 538
  identifying isotope patterns 569
  setting up protein ratios 500
  setting up quantification method 485
  specifying label modifications 488
  specifying quantification channel names 495
  specifying quantification channels 488–489, 506, 510
  summarizing settings for 523, 526
  troubleshooting 574
  See also Precursor Ions Quantifier node
precursor ions
  annotating spectra with 58
  displaying associated scan 348
  displaying charge state 263, 348
  displaying in Peptide Identification Details view 302–303, 307
  displaying isotope pattern 337
  displaying mass-to-charge ratio 263, 348
  displaying singly charged mass 263, 348
  in Peptide Identification Details view 311
  intensity 263, 348
Precursor Ions Area Detector node
  description 120, 698
  nodes not used with 483
  performing peak area calculation quantification 7, 103, 120, 483, 698
  unavailability for use in peak area calculation quantification 698
  using to access Quantification menu 468
Precursor Ions page 307
Precursor Ions Quantifier node
  description 120, 473, 699
  nodes not used with 473
  parameters 699
  performing precursor ion quantification 103, 120, 472–473, 699
  setting parameters 475, 698–699
  setting up quantification method 486
  summarizing node settings 523
  unavailability for use in peak area calculation quantification 699
  using to access Edit Quantification Method command 520
  using to access Quantification menu 468
Precursor Isotope Pattern view
  description 337
  displaying 337
  features 338
  opening 621
  parameters 340
Precursor m/z column 348
Precursor Mass peptide filter
  Mascot reports 374
  SEQUEST reports 370
  ZCore reports 378
Precursor MH+ column 348
precursor overtone peaks
  calculating mass window for 651
  removal by Non-Fragment Filter node 649
  removing 651, 655
  using Mass Window Offset 653
  using Remove Only Known Masses 654
precursor peaks 650
PrediSi database 453
PRF database 705
Probability column 293
probability score 671
Probability value
  Peptide Identification Details view 312
  Peptides page 233, 293, 353
Processing menu 22, 624
Processing Node Name column 241
Processing Node No. column 241, 292
Processing Setup icon 134
product ions
  intensity 335, 337
  mass-to-charge ratio 335, 337
product limitations 14
Protein Groups column 293
Protein Accessions column 293
protein annotation
  configuring 435
  creating workflow 437
  GO 433–434, 437
  Pfam 435, 437
protein binding GO Slim category 459
protein complex GO Slim category 463
protein databases 163
Protein Description Contains protein filter 364
Protein Descriptions column 293

Protein Group Members view
  description 285
  displaying 286, 402, 405, 620
  Is Master Protein column 406
  matching number of proteins displayed to # Proteins column 242, 402
Protein Identification Details dialog box 449
Protein Identification Details view
  copying colored bar on Sequence page 254
  Coverage page 246
  description 245
  displaying
    coverage 246, 255
    filtered-out peptides 254–255
    information from ProteinCenter 245
    ProteinCard 434
    PTMs 246, 253–255
    UniProt annotations of protein 249, 253, 255
  exporting Modification List page to Excel 254
  finding protein PTMs 253
  Modification List page 251, 256
  opening 619
  parameters 255
  protein sequence coverage bar 247, 256
  ProteinCard page 434
  PTM site localization scores 103
  PTM site probabilities 249, 255
  saving colored bar on Sequence page 254
  Sequence page 247, 256
  sequence table 248, 256
protein quantification 546, 548
Protein Quantification page 500, 513
protein ratios
  calculating 550
  calculating for multiconsensus reports 563
  calculating from peptide ratios 557
  defining peptide uniqueness 501
  displaying peptide ratios used to calculate 244
  displaying variability 500, 513
  setting up 497
  setting up peptide parameters used in 500
  variability used to calculate 561
protein references 164–165
Protein References of a Peptide view 287
Protein Score protein filter 365
protein sequences 164–165
protein uniqueness 468, 514
ProteinCard
  accessing 434, 449
  accessing data in ProteinCenter 449
  Biological Processes page 449, 456
  Cellular Components page 449, 455
  Diseases page 449, 458
  External Links page 449, 458
  Features page 449, 453
  General page 449–450
  Keys page 449, 452
  Molecular Functions page 449, 454
  Pfam identification number 435
  tabs in 434
ProteinCenter
  accessing annotation data in 438, 449
  description 2, 433
  displaying information from in Protein ID Details view 245
  retrieving annotations from GO and Pfam databases 437
  uploading search results to 446, 628
  Web server address 436
ProteinCenter page 448
Proteins column 293
proteins
  # Unique Peptides column 412
  accession keys 452
  annotation *See* protein annotation
  biological processes 456
  cellular components 455
  determining which to include in quantification 468
  diseases associated with 458
  displaying
    coverage 246, 255
    filtered-out 400, 630, 633, 635
    general information about 450
    group members 405
    peptides included 225
    PTMs 246, 255
    UniProt annotations 249, 253, 255
  elution time in Chromatogram view 333
  exporting to FASTA file 599
  filtering 387
    applying filters 387
    deactivating filters 392
    filters available 364
    removing filters 392
    result filters 379
    row filters 393
    specifying filters 361
  finding PTMs 253
  function of 454

proteins (continued)
grouping
algorithm used 407
Apply Strict Maximum Parsimony Principle 363
by algorithm in previous releases 412
by PSM confidence 363
by PSM Delta Cn 363
displaying other proteins belonging to same group 405
homologous 363
numbers on status bar 27
on Proteins or Peptides page shortcut menu 402–403, 630, 633
on Result Filters page 362, 402–403
on Search Input page shortcut menu 635
options 404
peptides with sequences not belonging to master protein 411
Protein Group Members view *See* Protein Group Members view
PSMs identified by multiple workflow nodes 412
ranking 402
sequence redundancy 401
turning off 406
grouping data by column header 227
groups in status bar 412
highest-scoring 221
homologous 404
master 402
master proteins 406, 411
members of groups 402
number displayed in MSF file 221
number of 402
Protein Identification Details view *See* Protein Identification Details view
PSM Ambiguity column 410
ranking 402
Report Item Distribution chart 257
retention time 333
retrieving information from ProteinCenter 433
scoring 359
selecting 232
Sequence Comparison view 283
sequence features 453
sequences 256
Web links to information 458
Proteins page
adding fixed row index numbers 231
adding row index numbers to visible and invisible rows 231
creating custom layouts 233
description 237
displaying 238
filtered-out rows 400
GO database information from ProteinCenter 433
master proteins 402
peptide groups 240
peptides included in protein 225
Pfam annotations 435
exporting search results to Excel 236
exporting to tab-delimited TXT file 15, 609
levels in 239
multiconsensus reports 353
numbering rows 230
parameters 241
Protein Identification Details view 245
Quan Info column colors 22
Ratio columns 495, 497
ratio count and variability 560
Ratio Count columns 560
removing row index numbers 232
Report Item Distribution chart 257
row filters 393
selecting peptides and proteins 232
Sequence Comparison view 283
Proteins page of the To CSV dialog box 611
Proteins page shortcut menu 630
Proteome Discoverer application
closing 62, 618
closing all windows 629
displaying
Help 629
installation guide 629
release notes 629
displaying user guide 629
displaying version information 629
features 2
filtering data 379
inputs 12
limitations 14
main window 19, 61
new features in latest release 14
opening 61
outputs 13
search wizards 63
system requirements xvi
workflow 9
Proteome Discoverer application icon 122
Proteome Discoverer application installation directories 717
Proteome Discoverer Daemon
calling from Xcalibur data system 132
Configuration page 129
connecting to remote server 130–131
connecting to server 122

Proteome Discoverer Daemon (continued)
creating parameter file for calling from Xcalibur data system 125
description 121
Job Queue page 127
monitoring job execution 127
output files
batch processing 147
preparing to run Proteome Discoverer Daemon 133
starting a workflow 124, 445
reannotating MSF files 444
Refresh icon 124
running on local server 154–155, 157
running on remote server 154–156
Start Jobs page 123–124, 128, 444
starting in window 121
starting on command line 153
starting workflow for batch and MudPIT processing 123
using as post-acquisition method 144
See also batch processing
See also MudPIT
Proteomics Standards Initiative 329
ProtXML files 13, 617
pRS Probability column 299, 691
pRS Score column 299, 691
pRS Site Probabilities column 299, 691, 694
PSM Ambiguity column 292, 410
PSMs column 242, 293
PSMs *See* peptides
PTM Analysis nodes 691
PTM analysis workflow 103
PTMs. *See* post-translational modifications
Public Files directory 147
PubMed database 455
purity correction factors
applying in Ratio Calculation page 498, 512, 555
iTRAQ methods including 488, 509
selecting in Quan Channels page 494, 510
using in reporter ion quantification 545
putative isoforms 694, 697

Q

Qual Browser 7, 205, 628
Quan Channel column 296
Quan Channels page
creating new quantification method 518
parameters on 506, 508
specifying label modifications for reporter ion quantification 492
specifying quantification channels for precusor ion quantification 489
Quan Info column
calculating peptide ratios 550
color coding 22
displaying peptide classification 548
including excluded peptide in quantification results 547
Peptides page 297, 535
Quan Spectra page 214, 217
Quan Usage column
including peptides in quan results 547
Peptides page 298, 533
QuanResultID column 295
Quantification Channel Values chart 529–530, 532, 623
quantification channels
displaying values 529
for precursor ion quantification 489
for reporter ion quantification 492
missing 498
setting up a quantification method 485
setting up for ratio reporting 495
with only one peak 498
Quantification menu 22, 468, 622
quantification method
adding 515–516
changing 520
checking the parameters set 502
deactivating 522
deleting 515, 521
exporting 515, 523
importing 515, 522
restoring original template 503
setting up for multiple MSF files 503
setting up for precursor ion quantification 485
setting up for reporter ion quantification 485
Quantification Method Editor dialog box
changing quantification method 520–521
Experimental Bias page 501
General page 504
loading multiple MSF files 503
opening 486, 622, 699, 702
opening from Quantification Methods view 515
parameters on 506
Protein Quantification page 500, 513
Quan Channels page 489, 492, 506, 508, 518
Ratio Calculation page 497, 511, 535
Ratio Reporting page 495, 510, 518
setting options for multiconsensus reports 566
setting up quantification method 486
Quantification Methods view
adding quantification method 516
changing quantification method 520
deactivating quantification method 522
exporting quantification method 523

Quantification Methods view (continued)
importing quantification method 522
opening 627
parameters on 515
removing quantification method 521
restoring quantification method template defaults 503
setting up quantification method 486
Quantification Spectrum chart 533, 538, 544, 624
Quantification Summary page
activating 351
description 498, 523
displaying 525
parameters for precursor ion quantification 526
parameters for reporter ion quantification 526, 528
quantification workflow
creating 103
See also precursor ion quantification
See also reporter ion quantification
Quantification workflow nodes 698
QuickGO browser 454
q-value 688, 690

R

Rank column 294
Ratio Calculation page 497, 511, 535
Ratio columns 244, 298, 495
Ratio Count columns
calculating ratio count 560–561
Protein Quantification page 500, 513
Proteins page 244, 560
ratio counts 500, 513, 561
Ratio Reporting page 495, 510, 518
Ratio Variability columns
calculating ratio count 560
Protein Quantification page 513
Proteins page 245, 560
Quantification Method Editor 500
Ratio Variability columns description 561
raw files
base peak chromatogram of
Chromatogram view 330
search wizards 72
contents of Xcalibur data system 12
determining charge state 75
displaying in Quantification Spectrum chart 544
displaying information about in Search Summary page 349
displaying name
Chromatogram view 333
Extracted Ion Chromatogram view 343
Fragment Match Spectrum view 337
job queue 28
Peptide Consensus view 322
Peptide Identification Details view 315
Precursor Isotope Pattern view 340
Quantification Channel Values chart 532
Spectrum view 335
exporting spectra extracted from 578–579
importing from BioWorks 720
in parallel workflows 105–106
input file type 63, 735
location required to generate Extracted Ion Chromatogram view 341
passing for Qual Browser operations 205
performing multiple searches on multiple 28, 121
processing multiple from multiple samples 91, 121
processing multiple from one sample 91, 101, 121
processing one from one sample 91
processing synchronously in Xcalibur data system 146
search wizards
displaying base peak chromatogram 84
displaying name in 84
excluding first and last minutes of data in 73
grouping spectra 75, 85
specifying multiple files in 72
specifying name in 72
selecting in Workflow Editor 93
specifying in Spectrum Files node 117, 638
specifying instrument that produced 78
specifying mass analyzer used to create 347
submitting multiple files to Workflow Editor 89
Rawfile and Scan Range Selection page 72, 84, 735
RDB equivalents 205
reannotation
MSF file 444
Proteome Discoverer Daemon 444
Re-Annotation node
description 16, 120, 443, 704
parameters 704
retrieving protein annotations 249, 253, 443, 704
receptor activity GO Slim category 459
Redundant peptide status 297
Refresh icon 51, 124
regulation of biological process GO Slim category 465
relaxed target false discovery rate
Peptide Validator node 681
Percolator 690
Release Notes command 629
Remove FASTA indexes confirmation box 190
Remove icon 51, 199, 216
replicates
calculating protein ratios in multiconsensus reports 563, 566
definition 560
in mixed mode 567

replicates (continued)
ratio counts 561
treating quantification results as 217, 220, 356, 564
variability 561
variability in multiconsensus reports 563
Report Item Distribution chart
description 257
displaying ToolTips 259
features of 259
opening 619
parameters on 274
parameters on Scatter Plots page 261
pie charts 278
Pie Charts page parameters 280
scatter plots 258
Venn diagrams 272
Report Item Distribution dialog box 258
reporter ion quantification
calculating peptide ratios 550, 553
checking the quantification method 502
co-isolation 513, 548
columns on Proteins page unique to 244
controlling protein and peptide ratios 497
correcting experimental bias 501
correction for isotopic impurities 498
creating a workflow for 477
creating workflow for 482
default methods available in 488, 493, 509
description 7, 476
displaying quantification channel values 529
displaying Quantification Spectrum chart 533
isotopic distribution values 545
missing reporter ions 535
performing 476
performing on HCD and CID scans 481
setting up protein ratios 500
setting up quantification method 485
specifying label modifications 488
specifying mass tags 493, 495
specifying quantification channels 492, 508
summarizing settings for 523, 526
troubleshooting 571
See also iTRAQ quantification
See also Reporter Ions Quantifier node
See also TMT quantification
Reporter Ions Quantifier node
creating a workflow 478
description 120, 699
nodes not used with 478
parameters 699
performing TMT quantification on HCD and CID scans 481
reporter ion quantification 103, 120, 699
setting parameters 479, 700
setting up quantification method 486
summarizing node settings 523
using to access Edit Quantification Method command 520
using to access Quantification menu 468
reproduction 465
Residues column 161
Restore FASTA indexes confirmation box 191
result filters 379
Result Filters page
displaying 360
filtering data 361
filtering data in MSF file 379
filtering data in MSF file before opening 218
filtering results 379
interpreting search results 221
limiting number of proteins displayed 218
purpose 359
results report *See* MSF files
Resume icon 51
retention time
exporting exclusion mass lists 602
for peptides on Search Input page 348
for precursor ions on Peptide Identification Details view 312
for proteins on Chromatogram view 333
ribosome GO Slim category 463
RNA 450–451, 460, 463
RNA binding GO Slim category 460
row filter menu 397–398
row filters
clearing all 398
deleting individual 398
filtering search results 379
icons 395
icons defined 395
purpose 354
row index numbers 230–231, 631, 633
Row Numbers command 631, 635, 633
RT column 298, 348
Run Sequence dialog box 138, 145, 153
Run Sequence icon 138, 145

S

sample fractions 101
Save a Parameter File dialog box 126
Save As command 637
Save As Template command/icon 26, 625
Save Filter Set dialog box 361, 389
Save Modification List as Microsoft Excel File dialog box 254

Save Processing Workflow Template dialog box 82, 89, 98
Save Report command/icon 24, 616
Scan Event Filter node
  description 118, 659
  parameters 659
  used for Mascot or SEQUEST analysis 98
Scan Extraction Parameters page 73, 84, 735
Scan(s) column 348
scatter plots
  clearing single data point 261
  clearing visible data points 261
  copying 260
  description 258
  displaying 258
  example 259
  parameters for 261
  saving 260
  selecting single data point 261
  selecting visible data points 260
  zooming 260
Score column 241
Score versus Charge State peptide filter 370, 426, 429, 431
Search Description page 81, 88, 737
search engine rank 387
search engines available 3
Search ID column 241, 292, 347
Search Input page
  adding fixed row index numbers 231
  adding row index numbers to visible and invisible rows 231
  creating custom layouts 233
  description 346
  displaying 346, 630, 632
  displaying filtered-out rows 400
  exporting search results to Microsoft Excel 236
  exporting to tab-delimited TXT file 15, 609
  immediately displaying peptides 225
  numbering rows 230
  parameters on 347
  removing row index numbers 232
  row filters 393
  selecting peptides and proteins 232
Search Input page of the To CSV dialog box 611
Search Input page shortcut menu 635
Search Report menu
  commands on 237, 619
  description 22
  making commands available 213
search results window
  closing 35
  hiding 33
  icons table 35
  moving 34
  using navigation icons 33
Search Summary page
  description 349
  displaying 349
  parameters 351
search wizards
  configuring search parameters 64
  FASTA files used 63
  Mascot. *See* Mascot search engine
  SEQUEST. *See* SEQUEST search engine
  spectrum source files used 63
  starting searches 71
  workflow involved in using 64
  ZCore. *See* ZCore search engine
Seattle Proteome Center 12, 580
sector field mass spectrometer
  specifying in Reporter Ions Quantifier node 480, 702
  specifying in Scan Event Filter node 661
  specifying in Spectrum Selector node 642, 646
Select Analysis File(s) dialog box 93, 102, 638
Select Modifications page 79
Selected Ion Monitoring (SIM) scan type 644, 662
Sender ID column 347
Sequence column 292
Sequence Comparison view
  description 283
  displaying 283
  features of 284
  opening 621
  parameters 285
Sequence page 247
Sequence Setup icon 137
Sequences column 161
SEQUEST node
  description 119, 671
  parameters 671
SEQUEST search engine
  Completing the *Wizard_name* Search Wizard page 81, 89
  configuring 65
  creating FASTA index 17, 185
  description 3–4
  highest-scoring peptide 224
  main score 345
  Modifications page 88
  options for calculating FDR 426, 429
  peptide filters 366
  Rawfile and Scan Range Selection page 72, 84, 735
  Scan Extraction Parameters page 73, 84, 735
  Search Description page 81, 88, 737
  Select Modifications page 79

SEQUEST search engine (continued)
SEQUEST Search page 76, 86, 736
starting 71
using template for data migration 721
Welcome to the SEQUEST Search Wizard page 71, 83
wizard 624, 735
XCorr value
applying filters 357
creating a report 353
creating custom layouts 233
Peptide Identification Details view 312
Peptides page 293
SEQUEST Search page 76, 86, 736
SEQUEST Search Parameters page
SEQUEST
SEQUEST Search Parameters page 723
shortcut keys 44, 46
shortcut menus 23, 630
Show 'Group By Column' Panel command 632, 634, 636
Show Chromatogram View command/icon 25, 330, 620
Show Distribution Chart
command/icon 24, 258, 272, 278, 619
Show Extracted Ion Chromatogram
command/icon 25, 331, 341, 621
Show Filtered Out Rows command 288, 630, 633, 635
Show Fragment Match Spectrum
command/icon 25, 336, 620
Show in Peptide View command 225, 630
Show in Search Input View command 630, 632
Show Isotope Pattern command/icon 25, 621
Show Job Queue command/icon 26, 28, 626
Show Legend command 637
Show Peptide Consensus View command/icon
displaying Peptide Consensus view 316
displaying post-transitional modification info 327
interpreting Dual-Peptide Consensus view 325
Search Report menu 619
toolbar icons 25
Show Peptide Groups command
color coding 21
deselecting peptide grouping 290
grouping peptides 413
grouping peptides by sequence 228
Peptides page 288, 632
Proteins page 630
Search Input page 635
Show Peptide ID Details command/icon 24, 619
Show Peptide Ratios command/icon 25, 550, 622
Show Peptides Group Members View command/icon 344
Show Peptides Group Members view command/icon 25
Show Peptides Group Members view/icon 620
Show Position Tooltips command 637
Show Precursor Isotope Pattern command/icon 338
Show Protein Group Members command 402
Show Protein Group Members view command/icon 25
Show Protein Group Members view/icon 620
Show Protein ID Details command/icon 24, 619
Show Protein/Peptide ID Details
command/icon 253, 300, 449
Show Proteins Covered by This Set of Peptides command/
icon 245, 405
Show Quan Channel Values command/icon 25, 529, 623
Show Quantification Spectrum command/icon 25, 533, 624
Show Sequence Comparison command/icon 25, 284, 621
Show Spectrum command/icon 25, 334, 620
signal transducer activity GO Slim category 460
SILAC 2plex (Arg10, Lys6) quantification method
description 469
precursor ion quantification 469
selecting in Quantification Editor dialog box 487, 490
selecting in Quantification Method Editor dialog
box 507
troubleshooting precursor ion quantification 575
SILAC 2plex (Arg10, Lys8) quantification method
description 469
precursor ion quantification 469
selecting in Quantification Editor dialog box 487, 490
selecting in Quantification Method Editor dialog
box 507
troubleshooting precursor ion quantification 575
SILAC 2plex (Ile6) quantification method
description 469
precursor ion quantification 469
selecting in Quantification Editor dialog box 487, 490
selecting in Quantification Method Editor dialog
box 507
troubleshooting precursor ion quantification 575
SILAC 3plex (Arg6, Lys4|Arg10, Lys8) quantification
method
description 472
precursor ion quantification 469
selecting in Quantification Editor dialog box 487, 490
selecting in Quantification Method Editor dialog
box 507
troubleshooting precursor ion quantification 575
SILAC 3plex (Arg6, Lys6|Arg10, Lys8) quantification
method
description 472
precursor ion quantification 469
selecting in Quantification Method Editor dialog
box 487, 490, 507
troubleshooting precursor ion quantification 575
SILAC *See* precursor ion quantification
Single Peak Channels column 295

single quadrupole mass spectrometer 480, 661
  specifying in Reporter Ions Quantifier node 702
  specifying in Spectrum Selector node 642, 646
Single Reaction Monitoring (SRM) scan type 644, 662
single-search reports 562
Size column 161
Spectrum And Feature Retrieval nodes 639
Spectrum Confidence Filter node 16, 119
Spectrum Exporter node
  description 120, 702
  exporting spectra 114
  parameters 702
  using with Spectrum Selector or Spectrum Grouper node 702
Spectrum File column 299, 348
Spectrum Files node
  creating a search workflow 92
  description 117, 638
  parameters 638
  peak area calculation quantification 483–484
  precursor ion quantification 472, 474
  reporter ion quantification 477–478
Spectrum Filters workflow nodes 659
Spectrum Grouper node
  description 118, 655
  parameters 656
  used with Spectrum Normalizer node 658
  using with Spectrum Exporter node 702
Spectrum Normalizer node 118, 658
Spectrum Processing workflow nodes 647
Spectrum Properties Filter node
  description 119, 662
  parameters 662
Spectrum Score Filter node 664
Spectrum Selector node
  creating a search workflow 93
  description 117, 640
  interpreting results 333
  peak area calculation quantification 483–484
  precursor ion quantification 473–474
  reporter ion quantification 477–478
  Search Input page 346
  selecting precursor mass to use 94
  using with Spectrum Exporter node 702
Spectrum view
  description 333
  displaying 334
  opening 620
  parameters 335
  zooming 334
spliceosome GO Slim category 462
SpScore column 293
SRF File Import Wizard 616
  Check Modifications page 722
  Complete the Wizard page 725
  converting BioWorks data 720
  File Selection page 721
  importing SRF files 222
  Operation done page 724
  Operation status page 724
  SEQUEST Search Parameters page 723
  SRF Import Description page 724
  Welcome to the SRF File Import Wizard page 721
SRF File Selector node
  creating a search workflow 92
  description 117, 638
  parameters 638
  reporter ion quantification 477
SRF files
  corrupted 738
  creating from OUT and FASTA files 730
  importing 222
  importing through SRF File Import Wizard 12, 721
  issues in importing 740
  selecting for data migration 719
  specifying with SRF File Selector node 117, 638
  verifying completion and usability 733
SRF Import Description page 724
SRF Importer utility 715
SRF Importer view 715
Start Jobs page 123–124, 128, 444
Start Mascot Search Wizard command/icon 25, 625
Start SEQUEST Search Wizard command/icon 25, 624
Start Workflow command/icon 26, 626
Start ZCore Search Wizard command/icon 26, 625
static modifications
  definition 196
  selecting in SEQUEST wizard 79
  setting for FASTA indexes 187, 194
  setting in SEQUEST 669, 675
  setting in ZCore 680
status bar 27, 412–413
Status column 161
strict target false discovery rate
  Peptide Validator node 681
  Percolator node 690
structural molecule activity GO Slim category 460
Substitution column 199
survey link xviii
SwissProt database 705–708
system requirements xvi

T

tab-delimited TXT files 13, 577, 609, 612, 617
tandem mass tag quantification. *See* TMT quantification
target rate 427–428
temporary FASTA files 164
Thermo Xcalibur Home page 134, 137
Thermo Xcalibur Processing Setup - Programs dialog box 135–136
Thermo Xcalibur Processing Setup dialog box 135
Thermo Xcalibur Sequence Setup dialog box 137
third-party nodes 95
TIC. *See* total ion current (TIC) chromatogram
TIFF format 577, 598, 637
time of flight mass spectrometer
  specifying in Reporter Ions Quantifier node 480, 702
  specifying in Scan Event Filter node 661
  specifying in Spectrum Selector node 642, 646
Tmap database 453
TMT 2plex quantification method
  as default 7, 476, 488
  selecting in Quantification Method Editor dialog box 493, 509
TMT 6plex quantification method
  as default 7, 476, 488
  selecting in Quantification Method Editor dialog box 493, 509
TMT quantification
  creating a workflow for 477
  default methods available in 476, 488, 493, 509
  description 7, 476
  isotopic distribution values 545
  on PQD and HC scans 481
  performing 476
  performing on HCD and CID scans 481
  Reporter Ions Quantifier node 7, 120
  specifying label modifications 489
  See also reporter ion quantification
To CSV dialog box 612, 617
  opening 610
  Peptides page 611
  Proteins page 611
  Search Input page 611
To PepXML dialog box 598, 617
toolbar in Proteome Discoverer application
  customizing 36
  icons on 24
  removing toolbar icon 38
  repositioning icons 41
  restoring defaults 41
toolbar menu 36
toolbars
  adding to Customize dialog box 39
  customizing features 41
  deleting from Customize dialog box 40
  making visible in main toolbar 36
  removing from main toolbar 38
  renaming toolbars in Customize dialog box 40
Toolbars page
  customizing dialog box parameters 48
  customizing the job queue toolbar 51
  making toolbars visible 37
  renaming a job queue toolbar 52
Tools menu 22, 628
ToolTips
  customizing 43
  displaying 43
  displaying in Report Item Distribution chart 259
  displaying shortcut keys with 43
  hiding 43
Top *N* Peaks Filter node 118
  description 16, 658
  parameters 658
total ion current (TIC) chromatogram 340
Total Ions column 299
transcription regulator activity GO Slim category 460
translation regulator activity GO Slim category 460
transport GO Slim category 465
transporter activity GO Slim category 460
treatments
  calculating protein ratios in multiconsensus reports 563, 565
  definition 561
  in mixed mode 567
  treating quantification results as 565
  variability 561
  variability in multiconsensus reports 562
TrEMBL database 706–708
triple quadrupole mass spectrometer 480
  Scan Event Filter node 661
  specifying in Reporter Ions Quantifier node 702
  specifying in Spectrum Selector node 642, 646
TXT files *See* tab-delimited TXT files

U

U.S. National Library of Medicine (NLM) 455–457
Uncheck All command 631, 634, 636
Uncheck Selected command 631, 634, 636
Undo All Zoom/Pan command 637
UNIMOD
  importing chemical modifications 202
  login screen 327
  updating chemical modifications 195
  Web site 329

Unimod Access No. column 200
UNIMOD Web view 327–329
UniProt database
  accession key 452
  annotations in Protein Identification Details view 246, 255
  annotations in sequence table of Protein Identification Details view 250
  in PTM analysis 15
  retrieving annotations with Annotate or Re-Annotate node 249, 253, 437
  specifying in ProteinCard 453
UniProt NREF database 707
Unique Proteins protein filter 365
UniRef database 707
UniRef50 database 707
UniRef90 database 707
Universal Protein Resource consortium 707
uploading search results to ProteinCenter 446
Used peptide status 297
User Guide command 629
user interface of Proteome Discoverer application 19

V

vacuole GO Slim category 463
variability
  displaying 500
  in peptide ratios 244
  in protein ratios 513
  inherent in experiments 563
  multiconsensus reports 562
Variability columns 244
Venn diagrams
  copying 274
  description 272
  displaying 272
  saving 274

W

Welcome to the Search Wizard page 71
Welcome to the SEQUEST Search Wizard page 83
Welcome to the SRF File Import Wizard page 721
Window menu 22, 629
Workflow Editor
  aligning nodes 96, 438, 625
  annotation workflow 437
  attaching Peptide Validator node to search engine nodes 94
  attaching Percolatorr node to search engine nodes 94
  changing names and descriptions of workflow templates 113
  color of nodes in 94
  creating new search workflow 91, 99
  creating parallel workflows 105
  creating PTM analysis workflow 103
  creating quantification workflow 103
  creating workflow for MudPIT samples 101
  deleting workflow templates 112
  description 5
Export Workflow to XML command 626
exporting spectra in multiple formats 114
exporting workflow to XML file 626
features 2
Import Workflow command 626
importance of understanding nodes 89
importing from MSF or XML file 112
importing workflows
  in MGF format 114
  in MSF or XML format 626
  in MZDATA format 114
  in MZML format 114
  in MZXML format 114
incorrect parameter node settings 110
inputs 89
job queue. *See* job queue
joining nodes 95
nodes available in 117, 637
opening 89
  existing workflow 109, 625
  new workflow 625
organizing nodes 95
outdated nodes 109
panes 89
parameter filters 98
parameters available in 91
renumbering nodes 96, 438
renumbering nodes in order 626
saving workflow as template 98, 625
saving workflow as XML template 114
setting node parameters 96
setting up false discovery rates
  Peptide Validator node 419
  Percolator node 421
specifying raw file 93
starting a new search 89
starting workflow 99, 626
using third-party nodes 95
warning symbols 109
  Workflow Failures pane 110
  workspace definitions 6
Workflow Editor menu 22, 625
Workflow Failures pane 110
Workflow Nodes pane 6

workflow templates
changing names and descriptions of 113
deleting 112
for data migration 718
opening 109
saving 98, 625
Workspace pane 6

X

x fragment ions
definition 7, 302
displayed in Peptide Identification Details view 302, 311, 314
flag representing 318
location of charge 300
used for spectrum matching 676
Xcalibur data system
calling Proteome Discoverer Daemon 132
creating a parameter file 125
exclusion mass lists 346
export options 577
exporting exclusion mass lists 600
installing 717
Qual Browser 205
Xcalibur Sequence Editor 136
Xcalibur Setup dialog box 717
Xcalibur Setup wizard 717
XCorr column 293
XCorr value
combining filters in multiconsensus reports 357
creating a report 353
creating custom layouts 233
Peptide Identification Details view 312
Peptides page 293
XIC. *See* Extracted Ion Chromatogram
XML files
exporting data to 577
exporting search results in 15, 596
input to Proteome Discoverer application 12, 112
output by Proteome Discoverer application 13
template for data migration from BioWorks 719
XML template 114

Y

y* fragment ions 311, 713–714
y fragment ions
abstracting proton from precursor 7
definition 7, 302, 714
displayed in Peptide Consensus view 317
displayed in Peptide Identification Details view 302, 311, 314
flag representing 318
generated by CID 3
location of charge 300
used for spectrum matching 677

Z

z fragment ions
definition 7, 303, 714
displayed in Peptide Consensus view 317
displayed in Peptide Identification Details view 302, 311, 314
flag representing 318
generated by ETD and ECD 3
location of charge 300
used for spectrum matching 677
ZCore node
description 119, 677
parameters 677
ZCore search engine
available search engines 3
configuring 70
creating FASTA index 17, 185
description 3–4
highest-scoring peptide 224
main score 345
mass spectrometers used with wizard 63
options for calculating FDR 426, 429
peptide filters 375
Probability value 233, 293, 353, 312
starting 71
wizard 625
Zoom Out command 637